Computational understanding of visual interestingness beyond semantics: literature survey and analysis of covariates

Mihai Gabriel Constantin[1], University Politehnica of Bucharest, Romania
Miriam Redi[1], King's College London, United Kingdom
Gloria Zen[1], University of Trento, Italy
Bogdan Ionescu, University Politehnica of Bucharest, Romania

Understanding visual interestingness is a challenging task addressed by researchers in various disciplines ranging from humanities and psychology to, more recently, computer vision and multimedia. The rise of infographics and the visual information overload that we are facing today have given this task a crucial importance. Automatic systems are increasingly needed to help users navigate through the growing amount of visual information available, either on the web or our personal devices, for instance by selecting relevant and interesting content. Previous studies indicate that visual interest is highly related to concepts like arousal, unusualness or complexity, where these connections are found either based on psychological theories, user studies or computational approaches. However, the link between visual interestingness and other related concepts has been partially explored so far, for example by considering only a limited subset of covariates at a time. In this paper, we present a comprehensive survey on visual interestingness and related concepts, aiming to bring together works based on different approaches, highlighting controversies and identifying links which have not been fully investigated yet. Finally, we present some open questions that may be addressed in future works. Our work aims to support researchers interested in visual interestingness and related subjective or abstract concepts, providing an in-depth overlook at state-of-the-art theories in humanities and methods in computational approaches, as well as providing an extended list of datasets and evaluation metrics.

CCS Concepts: •**Information systems** → **Content analysis and feature selection; Information extraction; Image search; Video search;** *Recommender systems;*

Additional Key Words and Phrases: interestingness, affective value and emotions, aesthetic value, memorability, novelty, complexity, coping potential, visual composition and stylistic attributes, social interestingness, creativity, humour, urban perception, saliency

1. INTRODUCTION

The extent to which people find an image (or a video) interesting, i.e. *holding or catching the attention* [Stevenson 2010] is called visual *interestingness* [Silvia 2005]. Nowadays, we are facing an era where visual interestingness is of unprecedented importance. Billions of images and videos are generated and spread worldwide through social media and photo sharing platforms on a daily basis. Such a continuous flow of visual information necessarily reduces the attention span that users can dedicate to each piece of content [Romero et al. 2011]. To help navigate through huge amount of information, content delivery systems need to retrieve and recommend *interesting* items to users. In this context, developing effective methods that comprehend and predict media *interestingness* becomes crucial for user engagement and retention.

Multimedia and computer vision researchers, traditionally interested in modelling *tangible* visual properties such as objects and scenes [Zhou et al. 2014], have recently started studying more *intangible* properties of the audiovisual content, including visual interestingness, from an algorithmic perspective. However, as shown in the examples in Figure 1, defining interestingness can be non-trivial. Many new challenges arise when designing frameworks for automatic interestingness comprehension.

Modeling interestingness is inherently an interdisciplinary research work. Interestingness depends on human perception, a subject widely studied by humanities and social sciences. Computer scientists willing to study these subjects should work with

[1]Equal author contribution. Part of this work was funded by UEFISCDI, grant 2SOL/2017, SPIA-VA.

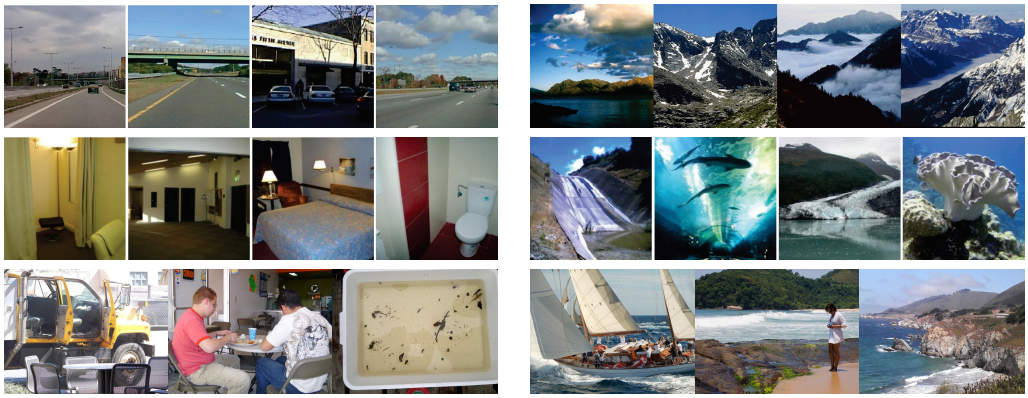

Fig. 1. Sample sequences of images rated with a (left) low and (right) high interestingness score. Images have been picked from: 8 scenes category dataset [Oliva and Torralba 2001], memorability dataset [Isola et al. 2011b], and visInterest dataset [Soleymani 2015].

experts in the field to gain a deeper understanding on the meaning of interestingness. Secondly, *collecting data* to train models for subjective perception requires the design of specific methodologies. As opposed to binary objective labels on objects and scenes, the collection of subjective judgments requires novel scales, interaction patterns and quality control mechanisms. Furthermore, in unsupervised cases, where the need for ground truth data collection is minimal, researchers would again need an interdisciplinary approach to mimic the human perception of interest or to find cross-domain correlations with other concepts. Thirdly, *creating features or learning systems* designed for interestingness rating or recognition becomes a completely new research task: when recognizing an object, e.g. a cat, a model creates representations of that object by minimizing differences between images of the same class; when recognizing interestingness, we want the model to enhance differences between images of the same semantic class and discover the differences between interesting cats and boring cats.

The algorithmic understanding of interestingness is, therefore, a non-trivial task. In this survey, we dive deep into the complexity of interestingness and compile an extensive but compact overview of related works tackling the understanding of interestingness and related concepts. While studying the literature that researches the notion of interestingness, it becomes clear that interestingness is not a stand-alone concept: many aspects of subjective perception such as emotions, aesthetics, memorability, are closely linked to interestingness. We summarize the works in humanities and social sciences dedicated to interestingness and related concepts. Then we approach methodologies to reliably collect data on interestingness and related concepts, providing an extended list of relevant datasets and evaluation metrics dealing with these concepts, allowing future researchers to further study the correlation between them. Finally, we look at computer vision and multimedia research works studying interestingness from a computational perspective.

While previous studies touched upon the concept of interest and other perceptual concepts (e.g., psychological interest [Silvia 2006], social interest and memorability [Amengual et al. 2017], visual attention and saliency [Zhao and Koch 2013], novelty detection [Markou and Singh 2003], etc.), no previous major survey dealt with the relation between interestingness and a wide variety of other perceptual concepts while analyzing the correlations with these concepts from the perspectives of social sciences, theoretical and experimental psychology and computer vision.

The rest of the paper is structured as follows. Section 2 investigates the definitions of the related concepts and proposes a taxonomy. An overview of the psychological aspects of these concepts is presented in Section 3, while Section 4 deals with the most relevant user studies and presents the main datasets and evaluation metrics. Computer vision approaches are shown in Section 5. Several applications are brought into discussion in Section 6, while Section 7 concludes the paper and formulates future open questions.

This work is intended for researchers and practitioners from various areas involved in or willing to explore the study of interestingness. This overview can be read at different levels of detail, thus suiting various level of expertise.

2. TAXONOMY AND DEFINITIONS

In this survey we will bring attention to a wide set of abstract and subjective concepts besides interestingness, either because these are deemed to be positively or negatively correlated with interestingness, or because the correlation between interestingness and these concepts is still largely unexplored. These concepts are listed in Table I and II. We selected these concepts according to the literature in various areas investigating the relation between interestingness and other aspects of visual perception.

Table I. Taxonomy. List of concepts mentioned or covered during our study in relation to interestingness. Semantically close concepts are grouped under their correspondent main theme.

	Theme	*Close concepts*
1	Interestingness	Interestingness
2	Affective Value and Emotions	Dimensional Emotion Space (Valence/Pleasantness, Arousal, Dominance) and Categorical Emotion Space (Happiness, Boredom, etc.)
3	Aesthetic Value	Aesthetic Value and Cuteness
4	Memorability	Memorability
5	Novelty	Novelty, Originality, Unusualness, Unexpectedness, Distinctiveness and Familiarity
6	Complexity	Complexity and Simplicity
7	Coping Potential	Coping Potential, Comprehensibility, Challenge and Uncertainty
8	Visual Composition and Stylistic Attributes	Symmetry, Balance/Harmony, Photographic Composition, Naturalness and Realism
9	Social Interestingness	Popularity and Virality
10	Creativity	Creativity
11	Humor	Humor, Irony and Sarcasm
12	Urban Perception	Urban Interestingness
13	Saliency	Saliency and Attention

Table I lists the concepts considered, grouped by a main common theme. For instance, popularity and virality are forms of social interestingness, thus, they come under this common main theme. In Table II, we highlight whether these concepts are deemed to be positively or negatively correlated to visual interestingness, according to user studies (see Section 4), psychological or cognitive theories (see Section 3), or based on some computational approaches (see Section 5); or whether the link between the two is still mostly unexplored. Correlation controversies are also highlighted and further discussed upon in the corresponding sections and indicate concepts that have been shown to be both positively and negatively correlated. These controversies may arise from different experimental setups, datasets used in the creation and testing of the prediction systems, differences in the questionnaires used in user studies or even

Table II. Taxonomy. List of abstract concepts considered in this survey and their respective correlations to interestingness. Correlations supported by (u) user studies or crowdsourcing, (t) psychological or cognitive theories or (c) computer vision approaches, are indicated with the corresponding letter. Some papers are presented as important examples of these correlations. Controversies, corresponding to concepts found to be both positively and negatively correlated with interestingness, are highlighted with *.

Positively correlated	*Negatively correlated*	*Mostly unexplored*
• Valence$^{(u,t)*}$ [Gygli et al. 2013] • Arousal$^{(u,c)}$ [Soleymani 2015] • Aesthetic Value$^{(u,t,c)*}$ [Hsieh et al. 2014] • Novelty$^{(u,t,c)}$ [Gygli et al. 2013] • Unusualness$^{(c)}$ [Zhao et al. 2011] • Unexpectedness$^{(t)}$ [Padmanabhan and Tuzhilin 1999] • Complexity$^{(u,t,c)}$ [Silvia 2005] • Coping potential$^{(u,t)*}$ [Silvia 2009] • Uncertainty$^{(t)}$ [Berlyne 1960a] • Balance/Harmony$^{(u,c)}$ [Jiang et al. 2013] • Naturalness$^{(u)}$ [Halonen et al. 2011] • Photo Composition$^{(c)}$ [Jiang et al. 2013] • Humor$^{(t,c)}$ [Jiang et al. 2013] • Urban interestingness$^{(u)}$ [Santani et al. 2017] • Saliency$^{(u)}$ [Elazary and Itti 2008] • Attention$^{(t)}$ [Berlyne 1971] • Popularity$^{(u,c)*}$ [Gygli and Soleymani 2016]	• Valence$^{(u,t)*}$ [Turner and Silvia 2006] • Boredom$^{(u,t)}$ [Finkenauer et al. 2002] • Aesthetic Value$^{(t)*}$ [Schmidhuber 2009] • Memorability$^{(u)}$ [Isola et al. 2014] • Coping potential$^{(u)*}$ [Soleymani 2015] • Challenge$^{(u)*}$ [Chen et al. 2001] • Virality$^{(u,c)}$ [Deza and Parikh 2015] • Popularity$^{(u,t)*}$ [Hsieh et al. 2014] • Familiarity$^{(u)}$ [Chu et al. 2013]	• Dominance • Cuteness • Originality • Distinctiveness • Comprehensibility • Symmetry • Realism • Irony, Sarcasm • Creativity • Urban Perception

the demographics of the subjects involved in the research process. The existing literature reveals that interestingness seems to be highly related to: high-arousal emotions, visual novelty, visual complexity, coping potential. Interestingness appears to be negatively related to: memorability, familiarity, low-arousal emotions (including boredom), social interestingness. These are to be explored in detail in Section 4. However, before doing so, we first provide a common understanding for each of the concepts via their definitions, as found in the relevant literature.

Interestingness. *Interestingness and Interest* — In one of his earlier works in this domain, [Berlyne 1949] views interest as a defining factor of human behaviour and motivation in particular and as a very relevant dimension in several branches of experimental psychology. While investigating some psychological concepts that influence interest [Berlyne 1970] points out that interest is created when new information is compared by a subject with information that already exists. [Chamaret et al. 2016] define visual interestingness as the quantification of an image's ability to induce interest in a user. From the observer's perspective, "interest" is indicated by the willingness to pay attention to a particular image. Interest, or "*situational interest*", is defined as "the appealing effect of an activity or learning task on an individual, rather than the *individual*'s personal preference for the activity" [Hidi and Anderson 1992], which [Silvia 2006] refers to as "interests". More broadly, interest is defined as an *emotion* related to exploration and learning [Silvia 2005]. [Silvia 2009] includes interest among the *knowledge emotions*, i.e. that family of feelings related to comprehension and thought. Figure 1 presents some examples of interesting and non-interesting image samples.

Affective Value and Emotions. Visual affective value is the ability of images and videos of conveying emotions, where "emotion" is any mental experience with high intensity and high hedonic content (pleasure/displeasure) [Cabanac 2002]. Emotions can be described in dimensional terms, such as valence or arousal, or through a categorical representations, such as happiness, anger, etc. *Dimensional Emotion Space: Valence Arousal Dominance* — The VAD model is the most widely used dimensional emotion space. Valence, pleasure or pleasantness is used to calculate the type of emotion - positive to negative; arousal refers to the intensity of an emotion; control or dominance

refers to the control of the subject over the emotion. Some more recent works suggest the existence of a fourth dimension, *novelty* or *unpredictability*, in the emotion space representation [Fontaine et al. 2007]. This fourth dimension may help to better represent emotions such as *surprise*, which refers "to a particular quality or dimension of emotional experience based on appraisal of novelty and unexpectedness" [Fontaine et al. 2007]. *Categorical Emotion Space* — [Ekman 1992] identified a universal facial expression set of 6 emotions, including "*anger*", "*disgust*", "*fear*", "*joy*", "*sadness*" and "*surprise*". [Plutchik 1980] considered 8 primary bipolar emotions, namely "*anger*" and "*fear*", "*joy*" and "*sadness*", "*anticipation*" and "*surprise*", and "*trust*" and "*disgust*". In Plutchik's wheel of emotions, this definition is extended to a set of 24 emotions, augmenting the sets of emotions along the intensity dimension. This extended set includes "*boredom*", "*distraction*", "*trust*" and "*interest*", where interest is considered as a mild form of "*anticipation*". *Boredom* — Boredom "is a feeling of displeasure due to a conflict between a need for intensive psychological activity and lack of stimulation or inability to be stimulated thereto" [Fenichel 1951]. Boredom has indeed been defined as "a state of relatively low arousal and dissatisfaction, which is attributed to an inadequately stimulating situation" [Mikulas and Vodanovich 1993].

Aesthetic Value. *Aesthetic Value* — Objects and images with aesthetic value are seen as "*beautiful*" or appealing, and elicit satisfaction, attraction and emotional well-being. Aesthetics is indeed that branch of philosophy studying the nature and human's perception of beauty in arts [Zangwill 2003]. Some works consider aesthetic value and pleasantness as two concepts which are strongly interlaced. For example, [Gygli et al. 2013] use image attributes such as "*aesthetic*", "*pleasant*" and "*expert photography*" as indicators for aesthetic. Appraisal of intrinsic pleasantness was constructed by averaging pleasantness and aesthetics scores in [Soleymani 2015]. However, we analyze herein the two concepts separately. *Cuteness* — Cuteness is a particular type of aesthetic appeal, specific to living beings, and related to a specific set of facial/body characteristics e.g. a "baby face", that induce the motivation in others to take care of it [Lorenz 1971].

Memorability. The memorability of an image is its intrinsic ability of being stored by our visual memory: it reflects the extent to which the image can be remembered by human mind [Isola et al. 2011b; Bylinskii et al. 2015b; Khosla et al. 2015].

Novelty. *Novelty and Familiarity* — Novelty is a concept quite familiar to people. The term "new" is used in day-to-day language. [Maher 2010] defines novelty as "a measure of the distance from other artifacts in the space". The opposite of novelty is *familiarity*. [Chu et al. 2013] described familiarity not only as the recollection of a certain event, but also as a form of "*associative recognition*", as a "sense of knowing that brings about meaning based on previous experiences". *Originality* — Novelty and originality are both related to the notion of *unusualness*, namely the rarity in the occurrence of a given event/artifact [Jackson and Messick 1965]. Novelty is the property of artifacts which have unusual attributes compared to artifacts of the same class, while originality refers to the rarity of the artifact as a *whole* [Verhaegen et al. 2012]. *Unexpectedness and Unusualness* — Unexpected is defined as something surprising or unforeseen. Something unexpected can also be unusual. Still, the subtle difference between unexpectedness and unusualness is that the former is mostly related with expectation and surprise [Roseman 1996]. *Distinctiveness* — Distinctiveness describes events that violate the prevailing context [Hunt and Worthen 2006]. It is related to events' uniqueness, which makes the event (or artifact) easily distinguishable (different) from all others of the same class.

Complexity. *Complexity and Simplicity* — Complexity may be defined as "the amount of variety and diversity in a stimulus pattern" [Berlyne 1960b]. The opposite of complexity is *simplicity*. [Berlyne 1960a] selected the following important properties

for this concept: complexity increases when the number elements increases, when the elements are dissimilar, and decreases when elements can be perceived as a group or unit. In general, the notion of complexity characterizes an artifact (in our case, some visual content) made out of many different parts. Such parts interact among themselves in many different ways, creating an element (emergence) greater that the sum of its parts [Anderson et al. 1972].

Coping Potential. *Coping Potential and Comprehensibility* — In the context of interestingness, coping potential is defined in [Silvia 2005] as the ability to understand a new and complex concept. Studies such as [Soleymani 2015] used a comprehensible-incomprehensible scale to determine the coping potential. *Challenge* — Challenge is defined as the difficulty level of a stimulus relative to subject's ability to understand it [Chen et al. 2001]. *Uncertainty* — Uncertainty is defined as the lack of information about an event, for example about whether, when, where, how, or why an event has occurred or will occur. It is an averse state that people are motivated to reduce. Uncertainty is inherently tied to the human ability of comprehension. To this reason, we include this concept herein.

Visual composition and Stylistic Attributes. We group the concepts of symmetry, balance, harmony and photographic composition under the visual composition theme; naturalness and realism can be defined as stylistic attributes. *Symmetry* — Symmetry reflects the extent to which shape and position of visual parts match after translation, rotation or mirroring with reference to an axis or a point in space [Lockwood and Macmillan 1978]. *Balance / Harmony* — Somehow related to symmetry, compositional balance "unifies the structural elements of a pictorial display into a cohesive organization or framework that helps determine the role of each element within a composition" [Locher et al. 2001]. *Photographic Composition* — The composition of a photograph reflects the way in which elements are organized in the scene space. In visual arts, visual composition refers to the ordering of elements such as lines, colors, textures, shapes in the photo frame [Grill and Scanlon 1990]. *Naturalness* — Naturalness is defined in [Janssen and Blommaert 2000] as "the degree of match between the internal representation of an image and memory". *Realism* — Realism is defined in [Ke et al. 2006] as the quality that differentiates an image to look rather "real" than "surreal". Indeed, professional photographers have the ability to make a picture look surreal, thus standing out of the context, by deploying a variety of techniques.

Social Interestingness. *Popularity* — Popularity is the quality of being liked or accepted by a large number of people. When referring to online content (e.g., stories, images, posts, etc.) this quality is usually measured by the number of views or likes. *Virality* — Virality indicates "the tendency of a content either to spread quickly within a community or to receive a great deal of attention by it" [Guerini et al. 2013].

Creativity. A creative artifact is something *new* (i.e. novel compared to artifact of the same class) and *valuable* (i.e. it has some affective or aesthetic benefit for its observer) [Weisberg 1999].

Humour. *Humour* — Humour is that quality of texts, images and any form of expression to generate laughter and provide entertainment. There are different types of humour: "Affiliative" is related to jokes, funny things or witty banters; "Self-Enhancing" is a coping mechanism; "Aggressive" is similar to sarcasm or derision, and "Self-Defeating" [Ruch 2008]. *Irony* — Irony is a literary form, based on the figure of the *simile*, used to emphasize meaning by using words or expressions that deliberately states the opposite of the truth or understates a factual connection [Muecke and Muecke 1969]. *Sarcasm* — Though sarcasm and irony are used interchangeably, sarcasm is a particular form of irony, generally involving malice, the desire to put someone

down, and it can be seen as *aggressive humour* [Martin et al. 2003] or "hostility disguised as humour"[2].

Urban Perception. The perception of urban environments reflects the way in which people sense and respond to city spaces around the world. Examples of urban responses are "*safety perception*", "*urban aesthetic judgement*" or "*urban interestingness*" [Santani et al. 2017].

Saliency. *Saliency* or salience "characterizes some parts of a scene, which could be objects or regions, that appear to an observer to stand out relative to their neighboring parts" [Borji and Itti 2013]. *Visual Attention* — Attention is defined in psychology as the "concentration of awareness on some phenomenon to the exclusion of other stimuli" [McCallum 1999].

3. HUMAN UNDERSTANDING OF INTERESTINGNESS AND COVARIATES

The human and subjective perception of visual data has been studied, extensively, in psychology and philosophy. In the following section we will analyze the meaning of interest, interestingness and covariates with reference to humanities studies: What do these concepts mean, and How do humans perceive visuals from a theoretical perspective? A summary of the main studies is presented with Table III.

Table III. A summary of the main theoretical studies on interestingness. The articles are grouped according to their perspective on interest: appraisal theory, motivational, physiological etc.

Source	*Short description*	*Main focus*
[Silvia 2005]	Interest from the perspective of appraisal theory: novelty and coping potential are the main components of this appraisal structure.	Appraisal theory
[Fredrickson 2001]	Interest as a driving force for long-term goal achievement.	Motivational
[Silvia 2006]	Interest as a motivator when dealing with some tasks; boredom as opposite of interest	Motivational
[Izard and Ackerman 2010]	Interest as a motivational element for learning and exploring (tied to survival)	Motivational
[Hess and Polt 1960]	Physiological reactions to different degrees of interesting stimuli.	Physiological
[Tomkins 1962]	Interest as a (positive) emotion associated with novelty, motivating people to explore new things	Emotional
[Silvia 2008]	Interest as a (positive) emotion, because of the physiological changes, expressions and subjectivity.	Emotional
[Berlyne 1960a]	Influence of several factors on human interest, including uncertainty, complexity and novelty.	Related concepts
[Padmanabhan and Tuzhilin 1999]	Unexpectedness as a component of human interest; unexpectedness impacts the subjectivity of interest.	Related concepts
[Schmidhuber 2009]	Links between interestingness, novelty, creativity and aesthetic appeal.	Related concepts

3.1. Interestingness

The extent to which people are interested in an image or a video is called visual interestingness. Interest is a very complex notion, and relationships between interest and other cognitive dimensions are not linear [Silvia 2009]. However, a substantial corpus of work from experimental and cognitive psychology have systematically investigated the meaning of interest and interestingness from a theoretical perspective.

What is interesting? — One of the first theories on interestingness [Berlyne 1960a, 1970] identifies a family of variables which determines whether something may result

[2]https://www.psychologytoday.com/blog/think-well/201206/think-sarcasm-is-funny-think-again

interesting or not to an observer: complexity, novelty, uncertainty and conflict. These variables indicate that interest arises in the human brain from a comparison between the incoming information and the existing knowledge within an observer. This influential theory highlights the subjective and dynamic nature of this attribute.

Interest as an Emotion — Interest can be described according to the theory of appraisals, which states that emotions are consequences of our evaluations (appraisals) of a given event. For example, [Silvia 2005] demonstrated that interest is generated by positive appraisals of novelty (is this novel?) and positive appraisals of coping potential (can I understand it?): something looks interesting if we find it new and easy to understand. [Tomkins 1962] was the first to view interest as a (positive) emotion, associated with novelty, which motives people to explore new things. Furthermore, [Hidi and Anderson 1992] found that interest can be caused by inherently emotional content such as sex and violence. The appraisal model of interest can explain why people can have different interest response to similar events (inter-subject variability) and why interest can change dynamically over time (intra-subject variability). Taking as example the visInterest dataset (see Table V), which comes with individual ratings from 20 viewers on each proposed image, the Cronbach Alpha α coefficient[3] obtained for interestingness is 0.96. The values obtained for the other attributes were: 0.94 (arousal), 0.88 (comprehensive), 0.96 (natural), 0.85 (familiar), 0.94 (quality), 0.97 (appealing), 0.88 (coherent), 0.98 (pleasing), 0.84 (complex), 0.97 (boring), 0.91 (easy-to-understand). This indicates that a strong inter-user agreement on interestingness judgment is possible. As reference, for random ratings (image ratings range from 1 to 7), α is -0.12, averaged over 1,000 random rating generation iterations.

Motivational Function of Interest — Several works in psychology highlight the motivational function of interest and its beneficial effects. [Izard and Ackerman 2010] suggest that "interest motivates exploration and learning, and guarantees the person's engagement in the environment. Survival and adaptation require such engagement". The motivational function of interest extends to activities that are not inherently interesting or appealing. Interest can bolster motivation to complete tasks that are boring and tedious [Silvia 2006]. Other theories suggest that interest motivates people to develop diverse experiences that can be helpful when unforeseen events occur. [Fredrickson 2001] proposes that interest, like other positive emotions, lacks short-term functions associated with survival. Instead, it serves long-term developmental goals: curiosity about the new and possible broaden experiences, attracts people to new possibilities.

Interestingness and Physical Reactions — From a physiological perspective, researchers found that pupillary movements are related to interest [Hess and Polt 1960]: the more interesting the stimuli, the larger the pupil dilatation.

3.2. Affective value and Emotions

Interest can be seen as an emotion, but How do images arouse emotions, and How do other emotions connect to visual interestingness?

Visual Emotions — The first studies on the emotional impact of images can be found in psychology. In particular, [Valdez and Mehrabian 1994] investigate the link between emotions and colors. Their systematic study evidences the strong and consistent effect of saturation, brightness, and hue on emotions along the three VAD dimensions. Out of the three dimensions, Valence and Arousal were found to be the most influential in studying visual emotion perception and were used more than the Dominance factor in subsequent research works. For example, in [Soleymani et al. 2008], a high correlation was found between image valence and smile/laughter reaction.

[3]It is a measure for evaluating inter-user agreement that estimates the reliability of a psychometric test.

Interestingness and Emotions — Interest is viewed by many researchers as an emotion, having all the basic characteristics of an affective response: physiological changes, facial expressions, cognitive appraisal, subjectivity, and an adaptive role across the lifespan [Silvia 2008]. Interest is indeed the *mild form* of anticipation in Plutchik's wheel of emotions. But is interest related to positive or negative emotions? [Silvia 2008] exposed differences between happiness and interest: the latter pushes towards new, not necessarily positive experiences, while the former pushes to stay attached to known sources of positive emotions. This intuition is further verified by the study of [Turner and Silvia 2006] proving that disturbing and less pleasant paintings elicit interest in a variety of subjects.

Boredom: Antonym of Interestingness — [Silvia 2006] cites the importance of understanding boredom in order to understand interestingness. In perception and psychology studies, "*boring*" has been often used as the opposite of "*interesting*" [Finkenauer et al. 2002]. However, although interestingness and boredom are inversely related, while interestingness is an activating factor, boredom is not the opposite/absence of interestingness, but a factor that limits the level of interest [Berlyne 1970]. For example, in Fowler's model [Fowler 1965] people don't seek interestingness, they avoid boredom.

3.3. Aesthetic Value

We will briefly summarize here a series of works from aesthetics philosophy on the nature of visual beauty and its impact on interestingness.

Visual Aesthetic Value — Common sense suggests that *"beauty is in the eye of the beholder"*, meaning that aesthetic judgments depend on the characteristics of the individuals. Despite beauty being very subjective, early aesthetic theory [Kant and Pluhar 1987] states that if humans were able to set aside personal interests and desires, they would achieve *disinterested contemplation*. For example, [Reber et al. 2004] introduced the idea that, in order to be considered beautiful, the observed item must have some features that will allow the human subjects to process it fluently, proprieties defined as: "goodness of form, symmetry and figure-ground contrast". Also, we use subject-specific aesthetic criteria to evaluate visual artworks: [Birkhoff 1933] concluded that the idea of giving aesthetic judgements to artifacts from different categories is not possible.

Interestingness and Aesthetics — Similar to interestingness, aesthetic reactions can be physically measured by looking at pupil dilation (pleasant stimuli generate greater pupil dilations, while unpleasant ones tend to cause pupil constriction [Blackburn and Schirillo 2012]) However, whether aesthetic value and interestingness are positive or negatively related remains, today, an unanswered question. A theory relating interestingness and aesthetics comes from computational creativity literature: according to this theory, aesthetic is related to beauty provided it comes with novelty [Schmidhuber 2009] *"interestingness is the first derivative of beauty: What is beautiful is not necessarily interesting. A beautiful thing is interesting only as long as it is new, that is, as long as the algorithmic regularity that makes it simple has not yet been fully assimilated by the adaptive observer who is still learning to compress the data better"*.

Cuteness: a Particular Aesthetic Value — The cuteness "baby schema" (Cute subjects show facial/body characteristics such as round face and big eyes was first discovered by ethologist Konrad Lorenz [Lorenz 1971]). Unlike aesthetic appeal, cuteness is believed to be a "universal" judgement: it follows evolutionary adaptations which pushed parents to take care of children, thus ensuring the survival of the species [Lorenz and Wilson 2002]. Later studies confirmed that cuteness preferences tend to be homogeneous across cultures [Van Duuren et al. 2003].

3.4. Memorability

The mechanisms of visual memory have been widely investigated in the neuroscience field. Visual memory stores visual inputs in 3 different ways: a detailed, *long-term* sensory representation including spatial information, a *semantic* description, and a *schematic* visual structure [Phillips 1974]. Early studies on long-term memorability of visual data such as images and videos started in the late '60s, with contributions of many different fields, from psychology to computer vision. All of these studies came to the surprising conclusion that the human memory for multimedia data is actually much more accurate than expected and much more consistent over time.

Human Visual Memory Capabilities — [Shepard 1967] were among the first to demonstrate the massive storage capacity of human long term memory in relation to images. They compared visual data memorability to words and sentence memorability, concluding that memory capacity for images was higher than that for words and sentences. But do humans remember only the *gist* (a very basic idea of a certain scene) of what they saw or does memory have a high capacity for a more detailed description? [Brady et al. 2008] proved that human memory capacity goes beyond mere scene gist recognition, and subsequent works [Rensink et al. 1997] proved that, when subjects were intentionally trying to encode and memorize visual details, they would succeed in doing so with great accuracy. While memory capabilities are fundamental to store visual information, the content of the object to be memorized plays an equally important role. Several studies have concluded that memorability "is an intrinsic property of images" [Bylinskii et al. 2015b; Isola et al. 2011a,b, 2014].

Time and Visual Memory — Time plays an important role in visual memory capabilities. Some experimental studies [Rensink et al. 1997] came to the conclusion that in order to perceive either minor or major changes in the scene's details, the attention span (i.e. the time spent to encode scene details) is very important. Another interesting finding of [Vogt and Magnussen 2007] was that participant's response time to memory tests was inversely proportional to the time span between the first phase of the test, i.e. image memorization, and the second phase: the actual recognition test. Two possible explanations come up from this finding: (1) Older memories might be better consolidated (and this would mean that they are more accessible). (2) In time, the scene memory forgets redundant parts of the initial information and the data quantity necessary for in-memory search decreases, speeding up the decision.

Interestingness and Memorability — In Section 4, we listed existing user studies on memorability, interestingness in the context of images, showing a negative relation between these 2 concepts. Opposite to these findings, in the context of textual data, few works in psychology and neuroscience have investigated the links between memory and a direct or indirect form of interest (personal relevance), showing that, in the elderly, interest plays an important role for remembering pieces of textual content [Germain and Hess 2007].

3.5. Novelty

Existing studies of the concept of novelty have been mainly investigated its relation to other concepts explored by this survey.

Interestingness and Unexpectedness — [Padmanabhan and Tuzhilin 1999] identify unexpectedness as a subjective measure that indeed contributes to the subjectivity of interestingness. The subjectivity of this concept is due to the fact that every person builds their own expectation, thus unexpectedness arises when these expectations are not met. In [Padmanabhan and Tuzhilin 1999], the second concept contributing to interestingness and its subjectivity is actionability, referring to the fact that a user is able to react to an unexpected pattern to his or her advantage. [Padmanabhan and

Tuzhilin 1999] define interestingness as a measure of how much it "shakes" the existing system of our belief, thus making unexpected patterns more interesting than the expected ones.

3.6. Complexity

The notion of complexity is intrinsically related to the notion of aesthetic beauty. One of the first research works studying ways to quantify complexity and aesthetic perception was made by [Birkhoff 1933]. Using a system based on psychological rewards he defined aesthetic measurement as a ratio between order (related to symmetry or harmony) and complexity.

Interestingness and Complexity — Complexity was often cited in psychological studies regarding interestingness, as early as the 1950s. The novelty-complexity appraisal structure [Silvia 2005] also indicated the importance of complex events in the understanding of human interest. The complexity of the visual composition has been found by different researchers to be highly related with interestingness [Berlyne 1960a; Grabner et al. 2013; Soleymani 2015]. Birkhoff's aesthetic measurement ratio [Birkhoff 1933] also seems to suggest a connection with the idea that images that are deemed too hard to comprehend - perhaps very high complexity factor - capture the subject's attention less, as theorized by [Silvia 2005].

3.7. Coping Potential

Coping potential was introduced while trying to define the appraisal structures that influence interestingness, first used by [Silvia 2005] and then further studied in [Soleymani 2015].

Interestingness and Coping Potential — In the theory of the appraisal structure of interest, [Silvia 2005] describes the two phases which lead to interest: appraisal of an event as "new, complex and unfamiliar (a high novelty-complexity appraisal) and as comprehensible (a high coping-potential appraisal)" [Silvia 2009].

Interestingness and Uncertainty — Uncertainty was one of the concepts described by [Berlyne 1960a] as having a strong connection with interestingness, and that, especially in situations where the stimulus creates a conflicting situation, uncertainty leads to anticipatory arousal for the subject.

3.8. Visual Composition and Stylistic Attributes

Visual composition is strongly connected to the affective reaction of the observer. For example, subjects that entirely fill the visual frame become more dramatic and cancel surrounding distractions. Also, many authors, artists and members of the media community tackled the importance of composition in creating an aesthetically pleasing image or videos [Krages 2012]. In particular, a number of compositional properties have been proven to affect psychological reaction to visual artistic forms.

Symmetry — Similar to the notion of complexity, symmetry has always been related to positive aesthetic preferences. However, unlike complexity, whose influence on aesthetic perception might vary from subject to subject, the regularities introduced by symmetric patterns universally elicit positive aesthetic responses [Cárdenas and Harris 2006]. Researchers of experimental psychology, in particular from the Gestalt School [Koffka 2013], have widely explored the importance of symmetry for visual perception: the human mind naturally perceives and combines visual (unconnected) elements as being symmetrical and arranging around a center point.

Naturalness — [Choi et al. 2009] found and analyzed four factors that influence perceived image naturalness: "colorfulness", "sharpness", "reproduction of shadow" and absence of "washed-out appearance".

Balance / Harmony — The overall visual balance generates the perception, although subjective, of visual stability and correctness [Lindell and Mueller 2011]. For example, experimental psychology studies showed that people tend to relate objects to the center of the frame, and prefer visual compositions reflecting the natural vertical arrangement of objects in space (e.g. a light bulb above the center and an aquarium bowl below the center) [Sammartino and Palmer 2012].

Photographic Composition — In photographic composition, colors also play a very important role in the compositional aesthetic value [Datta et al. 2006]: they have the power to change the affective value conveyed by artworks [Wexner 1954]: for example, in general, *red* is associated to excitement,*yellow* to cheerfulness, *blue* is related to trust [Wexner 1954] and *green* is seen as "fresh" and "pleasing" [Mahnke 1996].

3.9. Social Interestingness

Social interestingness is the amount of attention that an image receives due to social interactions. This quality has been investigated in contrast to visual interestingness [Hsieh et al. 2014]. While the latter can be explained purely by the visual content, the former needs further information such as the social context or the number of followers of the content publisher. [Hsieh et al. 2014] consider both virality and popularity as indicators of social interestingness, though these indicators are often considered conceptually different. Virality mostly refers to the probability of being reshared, thus allowing the content to spread quickly, while popularity refers to the probability of being liked.

Popularity — An item or person is defined as popular if it/she/he is able to attract collective admiration and interest. Interpersonal popularity can be categorized into two main notions: "*sociometric popularity* (being highly accepted or liked by peers)" [Asher and McDonald 2009] and *perceived popularity*, i.e. the popularity of a person as perceived by peers, linked to how visible, influencing and prestigious the person is known to be [Parkhurst and Hopmeyer 1998]. Image popularity largely depends on social ties within the network such as user popularity: number of contacts, number of groups and average users in groups [Khosla et al. 2014]. Indeed some images or videos can be highly rated by social media sites like Flickr and Pinterest, not because of their intrinsic interestingness but due to their social context.

Virality — A multimedia item "gone viral" is a special popular image or video that (1) has rapidly become popular, (2) similar to the contagion of infective viruses it had gained popularity through person-to-person spreading. Media studies have shown that viral videos tend to elicit strong affective response, especially related to anger [Nelson-Field et al. 2013] and that enjoyment, involvement, presence of celebrities, unlike brand category, are important for viral advertising videos [Southgate et al. 2010]. [Deza and Parikh 2015] studied the correlation between virality and a number of concepts including aesthetics, humor ("*funny*"), affective value ("*relaxed*", "*calm*", "*aggressive*"). Several of them were positively correlated, including funny, cartoonish, cute, scary etc., while others were negatively correlated like relaxed, serene, tired etc., showing that, in general, emotions with higher arousal were eliciting higher social interest.

3.10. Creativity

The study of the notion of creativity involves many disciplines, from psychology to theology, from economics to mathematics. [Rhodes 1961] defines four different research directions investigated by creativity researchers: "The person who creates"; "The cognitive processes involved in the creation of ideas"; "The environment in which creativity occurs or environmental influences"; "The product that results from creative activity". A vast number of psychology and neuroscience research works focused on *person, processes* or *environments*, namely the way in which human beings produce

creative artifacts. Since we are looking at works in computer science that detect creative traits of multimedia data, we are interested in defining the *creative product*, the human responses to creativity and the evaluation of creative products.

Defining "Creative" — Researchers have debated for nearly a century about the meaning of *creative*. [Barron and Harrington 1981] first proposed a definition involving the idea of usefulness and novelty. Today, researchers seem to agree that "All who study creativity agree that for something to be creative, it is not enough for it to be novel: it must have value, or be appropriate to the cognitive demands of the situation" [Weisberg 1999]. Not only creativity is related to imagination and novelty, but also a creative artifact must have a value, a benefit for the community or the audience it addresses. On the same line, [Maher 2010] investigated unexpectedness as one of the three essential criteria for creativity, together with novelty and value. The major difference between novelty and unexpectedness lies in the sequential nature of expectation.

Interestingness and Creativity — In his manuscript relating beauty, creativity, and interestingness, [Schmidhuber 2009] states that the curiosity of artists and scientists drives them to produce new, non-random, non-arbitrary creative artifacts. Such curiosity drive maximizes interestingness, which is, in its turn, "the first derivative of subjective beauty or compressibility".

3.11. Humour, Irony and Sarcasm

Humour — Humour is generally considered by psychologists a "healthy practice" that helps release tension and hostility [Firth 1957]. Among the most popular theories explaining humour, we can find the following: (1) The *"incongruity theory"*, stating that humour works by violating recipient's expectations [Ritchie 1999]; (2) The *"superiority theory"* associates humour with people's superiority compared to others' misfortunes [Mulder and Nijholt 2002]; (3) The *"benign-violation theory"* suggests that humour happens when "something seems wrong, unsettling, or threatening, but simultaneously seems okay, acceptable or safe" [McGraw et al. 2012]; (4) A *"general theory of verbal humour"* (GVLH) was developed by [Attardo and Raskin 1991]. It states that humour results from the activation of 2 different and incompatible scripts in a single text. Mainly related to cartoons and audiovisual productions, little work has been done on exploring humour in visuals and its psychological characteristics.

Irony and Sarcasm — While the notion of irony (sarcasm) has been widely investigated in its textual/verbal form, irony is, by nature, multimodal [Attardo et al. 2003]: ironic expressions involve audio (intonation pitch) and visual (facial expression) cues [Attardo et al. 2003]. Studies on multimodal texts, for example, showed that images help readers understand ironic cues. More specifically, while elements from image composition would not help with irony detection task, the facial expression of the speaker and the types of objects in the images are good ironic cues [Utsumi 1996]. An open questions remains: Can the combination of pictorial elements generate visual metaphors, which, in their turn, can create visual irony?

Interestingness and Humour — Although the impact of humour for interestingness has not been studied before, the addition of humour has been proven to enhance the interestingness of informative speeches [Gruner 1970]. Humour has also been found to be helpful to increase class interest during teaching [Powell and Andresen 1985].

3.12. Urban Perception

Urban perception studies answer research questions such as: How do people experience urban spaces? What do people remember about cities? Why do people perceive spaces are safe, lively or beautiful?

Interestingness and Urban Perception — Beyond urban space recognizability and liveliness, little work has been done on understanding what make urban spaces interesting for individuals. [Gygli et al. 2013] found that the presence of cultural/historical/outdoor places in images is a positive indicator of visual interestingness. A large number of computer science research works have focused on city "point of interest" recommendation, modeling, detection. However, these areas in the city represent the most "popular", touristic spaces, rather than the most "interesting".

3.13. Saliency

While Saliency has been studied mainly from a computational perspective, psychologists have long investigated the mutual effects between attention and interest.

Interestingness and Attention — Attention and interest are highly related. Attention plays an important role in interest arousal [Burnham 1908], and interest drives attention in visual perception [Rensink et al. 1997]. The reverse, however, is also true: interest is the key to elicit attention and foster learning [Shirey and Reynolds 1988]. Visual attention is commonly used to measure interest [Berlyne 1971; Silvia 2005]. Still, though interest clearly involves attention, these two concepts do not equate. For example, studies on text interestingness found that the interestingness of a text does not simply increase the amount of attention given to that text [Hidi 1995]. Other experiments show that people pay less attention to interesting texts with respect to uninteresting ones [McDaniel et al. 2000].

4. HUMAN JUDGEMENTS ON DATA

Having reviewed the relevant definitions of visual interestingness and related concepts as well as their psychological studies, we now explore how data supports quantitative studies on these topics. More specifically, in this section we investigate three key aspects: (1) How subjective human judgments regarding interestingness (and related concepts) on multimedia data are collected, and how these judgments are used to study the correlation between interestingness and other concepts; (2) What are the existing publicly available, annotated datasets, and how they are collected; and (3) What are the common evaluation metrics which would allow assessing the performance of a computational framework for automatic interestingness prediction.

4.1. User studies

Several studies have focused on collecting human judgments on visual interestingness and related topics. In this section, we will outline the main research works that investigated the notion of interestingness in relation to other subjective properties using quantitative methods (e.g. crowdsourcing) through user studies. A summary is reported in Table IV.

Interestingness and Pleasantness — [Chen et al. 2001] investigated the link between situational interest and five dimensional sources such as "instant enjoyment", "novelty", "challenge", "attention demand" and "exploration intention". The authors concluded that instant enjoyment was the most important factor for interest, while exploration intent and novelty also had a good influence. Another interesting observation is that enjoyment is enhanced by novelty and intent for exploration. In a study on memorability, interestingness and other factors, [Gygli et al. 2013] found that image attributes such as "*pleasant*", "*exciting*" and "*makes happy*" are highly related to interestingness; specifically, pleasantness was the third most correlated concept after assumed memorability ("Is this image memorable?") and aesthetics. Conversely, image attributes such as "makes sad" are negatively correlated. In a study on user responses to modern art images, [Fayn et al. 2015] found positive correlations both between interest and pleasure ratings and between interest and arousal ratings. In

Table IV. A summary of the main user studies on interestingness and its covariates, from psychology (top) and computational approaches (bottom).

Source	*Short description*
[Aitken 1974]	Investigating the link between *interestingness*, *enjoyment* and *complexity* based on human responses to polygons.
[Berlyne 1963]	A study on the time that people is willing to spend on *complex* and *interesting* patterns, rather than on *simple* and *enjoyable* ones.
[Cupchik and Gebotys 1990]	A study on human responses to paintings highlighting *complexity*, *novelty* and *meaningfulness* as underlying dimensions for interestingness.
[Russell and George 1990]	A user study on paintings that reveals no correlation between *interestingess* and *enjoyment*.
[Chen et al. 2001]	*Instant enjoyment*, *novelty*, *challenge*, *attention demand* and *exploration intention* as five sources of situational interest.
[Turner and Silvia 2006]	A study revealing that the dimensions of *novelty-complexity* and *coping potential* have high effects on interest as well as the lack of correlation between high *pleasantness* and *interestingness*.
[Elazary and Itti 2008]	Investigating the link between *saliency* and *interestingness* within an image.
[Fayn et al. 2015]	A study investigating the human responses to modern art images revealing that *visual interestingness* positively correlates with *arousal* and *pleasure*.
[Halonen et al. 2011]	Finding a high correlation between *naturalness* and *interestingness* in images.
[Gygli et al. 2013]	Collecting human responses on *visual interestingness* to investigate its link with *memorability* and other attributes like *aesthetic*, *pleasantness*, *arousing*, *exciting*.
[Hsieh et al. 2014]	Finding a high correlation between *visual interestingness* and *aesthetics*, and a low correlation between *social interestingness* and *aesthetics*.
[Isola et al. 2014]	Finding a positive correlation between *aesthetics* and *visual interestingness*.
[Soleymani 2015]	Collecting human responses on generic images for several attributes: *interestingness*, quality, *coping potential*, *naturalness*, *pleasantness*, *familiarity*, *arousal* and *complexity*. Finding that *arousal* is the most important attribute which explains interestingness, followed by intrinsic *pleasantness*, *quality* and *complexity*.
[Gygli and Soleymani 2016]	Investigating the link between *interestingness*, *aesthetics*, *arousal*, *valence* and *curiosity*, using human judgments recorded on animated GIFs.
[Santani et al. 2017]	Revealing a positive correlation between *urban interestingness* and attributes like *pretty*, *accessible* and *preserved*.

particular, pleasure was found to have a stronger correlation to interest with respect to arousal. An appraisal structure called "intrinsic pleasantness" was also constructed by [Soleymani 2015] as an average between pleasantness and aesthetic scores given by a crowdsourcing experiment, which was shown to have a positive influence on visual interestingness. A positive link between interestingness and valence was also found in [Gygli and Soleymani 2016]. However, not all studies came to the same conclusions. In another study on user responses to classical paintings, [Turner and Silvia 2006] showed that there was no link between a high pleasantness and interestingness, and that interesting paintings were also classified as disturbing by subjects. Similarly, in a study where 140 subjects (aged 16 to 80 years old) were asked to rank 15 paintings according to 7 aesthetic responses, [Russell and George 1990] show that rating of interest and enjoyment are uncorrelated. These reveal that the concepts are highly related to the application domain and data.

Interestingness and Arousal — In a user study on the link between interestingness and other concepts, [Soleymani 2015] found that arousal is the most important attribute which explains interestingness, followed by intrinsic pleasantness, quality and complexity. A positive correlation between visual interestingness and arousal was also found in user studies by [Fayn et al. 2015; Gygli and Soleymani 2016].

Visual Interestingness, Social Interestingness and Aesthetics — The user study conducted by [Hsieh et al. 2014] investigated the connection of visual interestingness, social interestingness (defined as social media popularity) and aesthetics. Visual interestingness and aesthetics of 200 images were assessed by a set of crowdsourced workers, each image being displayed to 10 workers. The participants were asked to rank the images with one of the five options: "*very boring*", "*boring*", "*neutral*", "*interesting*" and "*very interesting*". Regarding the social factor, the authors used scores extracted from the websites that hosted the images. The final results showed a high correlation between visual interestingness and aesthetics, and a low correlation between social interestingness and the other concepts. The important conclusion here is that images deemed by some social media sites as important are not necessarily beautiful or interesting from a purely visual point of view, even though these images are more frequently shared or up-voted by their users. However, in a different study based on GIF media interestingness, [Gygli and Soleymani 2016] found a positive correlation between social interestingness, expressed as popularity and number of likes, and visual interestingness. High correlation between interestingness and visual appeal is also found in [Gygli et al. 2013; Isola et al. 2014], where the authors show that interestingness and aesthetics subjective judgments are strongly correlated.

Interestingness and Memorability — Two studies [Isola et al. 2014] and [Gygli et al. 2013] explored the relation between memorability and interestingness using a crowdsourcing platform. [Isola et al. 2014] used a crowdsourcing game to collect memorability scores from 665 participants for 2,222 images. Out of this total batch, 30 participants also answered questions regarding their personal judgment about assumed memorability and interestingness ("Is this an interesting image?"). Using the same images, [Gygli et al. 2013] conducted "a binary experiment, where a user had to select the more interesting image from a pair". Final results demonstrated that assumed memorability is highly correlated to interestingness, while actual memorability is negatively related to interestingness.

Interestingness, Novelty and Familiarity — [Chen et al. 2001] analyze novelty as one of the five dimensional sources for situational interest (SI). Their study shows a direct influence of instant enjoyment on SI, with exploration and novelty having an indirect positive influence on SI via enhancing instant enjoyment. Similarly, the studies carried out by [Chu et al. 2013] on 42 participants have shown that images with unfamiliar context were considered more interesting than photos with familiar context. The opposite effect was found when dealing with human faces - self photos were the most interesting, followed by images of celebrities, friends and strangers.

Interestingness and Complexity — A traditional way in psychology to investigate human responses to visual complexity is by showing randomly generated polygons to people. A higher complexity is defined by a higher number of polygon edges. [Aitken 1974] conducted several experiments where people were asked to rank polygons according to how much "*interesting*" and "*enjoyable*" these were found. The research revealed that simple polygons are usually found more enjoyable, while highly complex polygons were usually rated as more interesting. In [Cupchik and Gebotys 1990], participants were shown a sequence of 12 to 15 paintings and they were asked to rate each of them in terms a number of dimensions, such as interestingness, aesthetic value, complexity etc. The analysis of results identifies complexity, novelty and meaningfulness as underlying dimensions of interest, while simplicity and emotional warmth as underlying dimensions for enjoyment. [Berlyne 1963] showed that people typically choose to view for a longer time interesting and complex patterns, rather than simple and enjoyable ones. Other studies, e.g. [Soleymani 2015], found a positive though weak correlation between complexity and interestingness.

Interestingness and Coping Potential — A coping potential appraisal structure was built by [Soleymani 2015], where participants were asked to rate images on three particular scales: "comprehensible-incomprehensible", "coherent-incoherent" and "easy to understand-hard to understand", and the experimental results on appraisal structures showed a negative effect on interest. Furthermore, the study shows that coping potential has a more negative effect on viewers' interest for the groups with higher openness traits, while complexity has a higher positive effect. This result indicates that people with higher openness are interested in stimuli which are more complex and difficult to understand. [Chen et al. 2001] studied challenge as one of the main factors of human interest. Challenge, however, had a small overall effect on situational interest and in some models a negative effect was detected, indicating that a high degree of challenge could diminish the interest in a certain event.

Interestingness and Naturalness — A high degree of correlation between interestingness and naturalness was observed by [Halonen et al. 2011] in an experiment with 24 participants who had to give a rating on two naturalness and interestingness scales. [Gygli et al. 2013] found a general preference for "outdoor-natural" scenes as opposed to man-made, indoor and enclosed spaces. [Soleymani 2015] found a positive though weak correlation between general interest and naturalness.

Interestingness and Urban Perception — [Santani et al. 2017] studied the correlation between six labels for a set of pictures depicting urban surroundings *accessible, dangerous, dirty, preserved, pretty* and *interesting*, asking both locals and online annotators to classify these images accordingly. The authors found a positive correlation between the accessible, preserved, pretty and interesting labels. While analyzing the differences between the two types of annotators, the authors noted that the online group tended to perceive images as more interesting, while the locals tended to perceive them as more dangerous and dirty.

Interestingness and Saliency — Visual interestingness and visual saliency are highly related. [Elazary and Itti 2008], for example, found that hot spots in image saliency maps can indicate interesting objects: the top salient location predicted by a bottom-up attention model matched with an interesting object in 43% of the cases, while in 76% of the cases the interesting object was overlapping with one of the top-3 salient areas.

4.2. Datasets

Data plays a critical role in exploring these concepts. In this section we report a description of the most relevant datasets on interestingness and related concepts, both in the case of images and videos or animated GIFs. These datasets were determined either to study the correlation between concepts or to experiment machine-based prediction approaches.

Image Datasets. The most relevant datasets are presented in Table V. [Gygli et al. 2013] considered two publicly available datasets annotated respectively with information on scene categories [Oliva and Torralba 2001] and memorability [Isola et al. 2011b] and extended them by crowdsourcing information on visual interestingness. The visInterest dataset [Soleymani 2015] was collected in order to investigate the link between visual interestingness and other factors such as complexity or familiarity. [Isola et al. 2011b] used the memorability game to measure image memorability for a set of 2,222 generic images sampled from the SUN dataset [Xiao et al. 2010]. A recent work [Khosla et al. 2015] proposed an efficient version of the memorability game, leading to the collection the largest dataset on memorability available so far, LaMem, consisting of 58,741 images.

The International Affective Picture System (IAPS) [Lang et al. 1999] is among the first datasets released to study attention and emotions. It comprises a set of 1,183 general photographs annotated with valence, arousal and dominance ratings. A subset of

Table V. Relevant image datasets for the analysis of interestingness and related concepts. For each dataset we indicate the number of images, labels, and the source of ground truth: human judgment is collected though crowdsourcing (C), from trusted annotators (T), or using information from social websites (W).

Dataset	*#Images*	*Label(s)*	*Source*
Scene categories, interestingness [Gygli et al. 2013]	2,688	interestingness, {*coast, mountain, forest, open country, street, inside city, tall building, highway*}	C
Memorability, interestingness [Gygli et al. 2013]	2,222	interestingness, memorability, attributes (e.g., aesthetic, pleasant, unusual, arousing, makes happy/sad, etc.)	C
visInterest [Soleymani 2015]	1,005	interestingness, quality, coping potential, naturalness, pleasantness, familiarity, arousal, complexity	C
LaMem [Khosla et al. 2015]	58,741	memorability	C
International Affective Picture System (IAPS) [Lang et al. 1999]	1,183	valence, arousal, dominance	T
IAPS (subset) [Mikels et al. 2005]	394	valence, arousal, dominance, {*amusement, anger, awe, contentment, disgust, excitement, fear, sadness*}	T
Art Photo (www.deviantart.com) [Machajdik and Hanbury 2010]	807	{*amusement, anger, awe, contentment, disgust, excitement, fear, sadness*}	W
Abstract paintings [Machajdik and Hanbury 2010]	228	{*amusement, anger, awe, contentment, disgust, excitement, fear, sadness*}	C
Emotion6 [Peng et al. 2015]	1,890	valence, arousal, anger, disgust, fear, joy, sadness, surprise, neutral	C
The Behance Artistic Media (BAM) [Wilber et al. 2017]	2.5M	{*scary, gloomy, happy, peaceful*}	C
DeviantArt; MART [Sartori et al. 2015b]	500;500	valence	C
15K Flickr [Schifanella et al. 2015]	15,686	beauty	C
Photo.net [Datta et al. 2006] (www.photo.net)	3,581	aesthetics, originality	W
CUHK [Ke et al. 2006] (www.dpchallenge.com)	12,000	{*high quality, low quality*}	W
Aesthetics (www.dpchallenge.com), Interestingness (www.flickr.com) [Dhar et al. 2011]	3,200; 8,000	aesthetic, social interestingness	W
AVA [Murray et al. 2012] (www.dpchallenge.com)	250K	aesthetic	W
Image Virality [Guerini et al. 2013] (plus.google.com)	174K	virality	W
Viral Images [Deza and Parikh 2015] (www.reddit.com)	10K	virality	W
Image Popularity [Khosla et al. 2014] (www.flickr.com)	2.3M	popularity	W

this dataset was further extended with the annotation of the main emotion elicited by the pictures [Mikels et al. 2005]. Along a similar line, [Machajdik and Hanbury 2010] released two additional datasets considering the same set of 8 emotional categories (i.e., *"amusement"*, *"anger"*, *"awe"*, *"contentment"*, *"disgust"*, *"excitement"*, *"fear"* and *"sad"*) for two sets of art photos and abstract paintings: 807 pictures from the art sharing website DeviantArt, and 228 abstract paintings which were rated though a web survey. [Peng et al. 2015] released a dataset of 1,890 images with annotations on valence, arousal and the 6 Ekman's emotions. For each image, they crowdsourced human judgments on the intensity of valence and arousal and on the dominant emotion. [Wilber et al. 2017] released the BAM dataset, which includes over 2.5M natural images and paintings with crowdsourced labels on the evoked emotion (scary, gloomy, happy and peaceful). [Sartori et al. 2015b] collected human judgements on the emotion (valence) evoked from abstract paintings for two abstract datasets. The use of abstract

art allows to purely focus on the effect of cues like colors or style, thus discarding the effect of high level content on the human emotional response.

Several works have exploited the information available on social photo-sharing platform such as Flickr to investigate concepts like emotions [Zhao et al. 2016], aesthetic judgement [Murray et al. 2012], social interestingness [Dhar et al. 2011], virality [Deza and Parikh 2015] and popularity [Khosla et al. 2014]. [Zhao et al. 2016] obtained personalized valence and arousal (VA) values of Flickr images based on users' comments and VAD norms of 13,915 English lemmas [Warriner et al. 2013]. [Datta et al. 2006] used aesthetic and originality ratings available on https://www.photo.net, a large photo sharing community. [Ke et al. 2006] obtained binary scores on image quality (high or low) on 12,000 images from http://www.dpchallenge.com by crawling 60,000 photos and selecting the top 10% and bottom 10% ranked out of these, based on the corresponding photo's average rating, where each photo was rated at least 100 times. [Murray et al. 2012] obtained aesthetic judgment scores of photographs using online human ratings from a photo-sharing website (http://www.dpchallenge.com). Furthermore, [Dhar et al. 2011] collected images from Flickr using Flickr's interestingness measure [Butterfield et al. 2014] which is based on the analysis of social interactions (e.g., popularity of the content and the owner, user patterns etc.). For this reason, we consider the Flickr score as a measure of social interestingness rather than visual interestingness. Specifically, it is related mostly to the concept of popularity, rather than virality, since social reactions like re-sharing are not taken into consideration in this case. Recently, [Schifanella et al. 2015] collected implicit beauty ratings for a large set of Flickr images, highlighting the mismatch between the value of social and visual interestingness in pictures from photo-sharing communities like Flickr. Image virality has been investigated analyzing public posts from Google+[4] [Guerini et al. 2013] and Reddit[5] [Deza and Parikh 2015]. In the first case, virality was measured based on the number of pluses, replies and reshares. In the second case, the virality score was computed based on the number of up and down votes and resubmissions received. [Khosla et al. 2014] explored image popularity using the number of views on Flickr.

Videos and GIFs Datasets. These datasets are summarized in Table VI. [Jiang et al. 2013] collected human judgement on video interestingness for 420 videos from 14 YouTube categories, e.g., "accessories", "clothing&shoes", etc. The assessors were shown pairs of videos and they were asked to rank the most interesting for each pair. Also, they collected 1,200 videos from Flickr by using Flickr interestingness-based ranking and 15 keywords, e.g., "basketball", "beach", etc. For each category, two sets of interesting and uninteresting videos were obtained by considering the top 10% and bottom 10% of interesting videos retrieved from each search on Flickr. [Gygli and Soleymani 2016] collected human judgement for a set of 2,739 animated GIFs, asking users to rate GIFs in terms of interestingness, aesthetic, arousal, valence and curiosity. [Redi et al. 2014] crowdsourced information on video creativity. They collected human judgement on creativity for 3,849 videos from Vine, asking users to indicate whether they find a video creative. Experimental results were reported for video subsets with 100%, 80% and 60% of user agreement. A dataset for the aesthetic assessment of videos, "*NHK - Where is beauty?*" [Takahashi and Sano 2013] was released for ACM MM grand challenge It consists of 1,000 broadcast videos, aesthetically ranked from the best (1) to the worst (1,000) based on human judgement. [Jiang et al. 2014] released the VideoEmotion dataset, a collection of 1,101 user generated videos downloaded from Youtube and Flickr based on the respective emotion tag. The videos were manually filtered to ensure that the assigned label matches the video emotional content. [Jou et al.

[4]https://plus.google.com/
[5]https://www.reddit.com/

Table VI. Relevant video and animated GIFs datasets for the analysis of interestingness and related concepts, where labels are provided as an overall score for the video/GIF. For each dataset we indicate the number of videos/GIFs, labels, and the source of ground truth: human judgment is collected though crowdsourcing (C), from trusted annotators (T), using information from social websites (W), or as unknown if this information is not available (U).

Dataset	*#Videos / GIFs*	*Label(s)*	*Source*
Youtube dataset [Jiang et al. 2013]	420	interestingness	T
gifInterest [Gygli and Soleymani 2016]	2,739	interest, aesthetic, arousal, valence, curiosity	C
Creativity in Micro-Videos [Redi et al. 2014]	3,849	creativity	C
NHK - Where is beauty? (ACM MM challenge, 2013) [Takahashi and Sano 2013]	1,000	aesthetic	U
VideoEmotion [Jiang et al. 2014]	1,101	emotions, {*anger, anticipation, disgust, fear, joy, sadness, surprise, trust*}	T
GIFGIF [Jou et al. 2014]	6,119	amusement, anger, contempt, disgust, embarassment, fear, guilt, happiness, pleasure, pride, relief, sadness, satisfaction, shame, surprise	C
LIRIS-ACCEDE [Baveye et al. 2015] (MediaEval 2016)	9,800	valence, arousal	C
Flickr dataset [Jiang et al. 2013] (www.flickr.com)	1,200	social interestingness	W
Movie Memorability Dataset [Cohendet et al. 2018]	660	memorability	T

2014] collected pairwise information for 6,119 animated GIFs on 17 emotions, and calculated the final scores for each emotion using the TrueSkill algorithm [Herbrich et al. 2006]. [Baveye et al. 2015] collected valence and arousal ratings for 9,800 video clips, each one lasting between 8 and 12 seconds. The videos were extracted from a set of 160 movies, including a wide variety of scenes spanning from violence, murders, landscape, interviews, positive scenes of daily life, etc. The annotation collection was done by applying a pair-wise protocol to crowdsourced annotations. [Cohendet et al. 2018] built a 660 video dataset for memorability prediction containing 10 second long videos extracted from Hollywood-like movies. Annotations were collected from 104 assessors, mainly researchers or students.

Table VII. Relevant video datasets for the analysis of interestingness within a video, where the labels were collected for video segments (or frames) extracted from videos. We indicate whether the labels were obtained by collecting human judgment either though crowdsourcing (C) or from trusted annotators (T).

Dataset	*#Videos*	*#Frames / #Segments*	*Label(s)*	*Source*
Webcam [Grabner et al. 2013]	20	3,180	interestingness	T
Yahoo Screen Video - Mouse Activity [Zen et al. 2016]	45	4,634	interestingness, mouse activity	C
Media Interestingness (MediaEval 2016) [Demarty et al. 2016]	52 train, 26 test	5,054 train, 2,342 test	interestingness	T
Media Interestingness (MediaEval 2017) [Demarty et al. 2017b]	78 train, 30 test	7,396 train, 2,435 test	interestingness	T

Video Datasets: predicting labels within a Video. The most relevant datasets are presented in Table VII. [Grabner et al. 2013] collected human judgement on visual interestingness for 20 webcam video streams, each one consisting of 159 images. A set of 46 participants to the study were asked to watch the image sequence, press a button when an interesting sequence started to be displayed and press it again to end the interestingness tag when the interesting part was over. Each sequence was viewed by at

least 20 persons. The overall interestingness rating of a sequence was computed as the average of the individual overall ratings. [Zen et al. 2016] crowdsourced information on users' mouse activity and video interestingness for generic videos of Yahoo Screen. The human judgment on interestingness was collected by asking the participants to watch the video and click a 'thumbs up' button when something interesting was shown and a 'thumbs down' button when something non-interesting was shown. Predicting visual interestingness was the focus of the MediaEval challenges that took place in 2016 [Demarty et al. 2016] and 2017 [Demarty et al. 2017b] which asked to identify from a given movie the corresponding key-frames, or video shots, that a common viewer would find as more interesting. The 2016 dataset consists of 5,054 shots and 5,054 key frames extracted from 52 movie trailers for training, and of 2,342 shots and 2,342 key frames extracted from 26 movie trailers for testing, while the 2017 dataset consists of 7,396 shots and 7,396 key frames extracted from 78 movie trailers for the training set, and 2,435 shots and 2,435 key frames extracted from 30 movie trailers for the testing set.

4.3. Evaluation metrics

Having in mind the possibility of designing a machine that would predict automatically the interestingness of multimedia, one has to think about assessing the performance. The very heterogeneous nature of data and the subjectivity of its annotations requires adapted metrics to highlight the performance of such a system. In this section we review the common metrics used in this particular context.

Correlation coefficients — In statistics, correlation coefficients are used to assess the dependence between two variables. Their range of values is from -1.0 to 1.0 and they measure the degree to which two variables are related: a correlation of 1.0 indicates a perfect positive correlation, a correlation of -1.0 indicates a perfect negative correlation. A correlation of 0 indicates that no relationship exists between two variables. The correlation value is usually reported with the associated p-value, which indicates the validity of the hypothesis of no correlation against the alternative of non-zero correlation. The latter is verified when p-value is low, such as less than 0.05. *Pearson's correlation*: This correlation coefficient is also called *Pearson product-moment correlation coefficient* (PPMCC) or *bivariate correlation*, and it is usually denoted as r or ρ. It measures the linear relationship between two variables and it is defined as the covariance of the two variables divided by the product of their standard deviations:

$$Pearson's\ \rho_{X,Y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y} \tag{1}$$

Spearman's correlation: This correlation coefficient is also called *Spearman's rank correlation* and it is usually denoted as r_s or ρ. While Pearson's correlation assesses the linear relationship between two variables, Spearman's correlation measures how well a monotonic function can describe the relationship between two variables. It uses the same formulation as Pearson's ρ, while the ranked values of the variables' observations are used:

$$Spearman's\ r_s = \rho_{rg_X, rg_Y} = \frac{cov(rg_X, rg_Y)}{\sigma_{rg_X} \sigma_{rg_Y}} \tag{2}$$

Kendall rank correlation: This correlation coefficient is also called *Kendall's tau coefficient* and it is usually denoted as τ. It measures the relationship between rankings of different ordinal variables:

$$Kendall's\ \tau = \frac{(N_C - N_D)}{n(n-1)/2} \tag{3}$$

where N_C and N_D is respectively the number of concordant and discordant pairs. Kendall's τ and Spearman's ρ can be seen as particular cases of a general correlation coefficient. With respect to Pearson's ρ, these two measures are more robust in the case of nonlinear relationships.

Root Mean Squared Error — RMSE is frequently used to measure the difference between the set of predicted values $\hat{y}$ and those actually observed y. It represents the standard deviation of the differences between the predicted and observed values:

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n}(\hat{y}_t - y_t)^2}{n}} \tag{4}$$

Information Retrieval standard metrics — Evaluation measures such as *Average Precision (AP)* and *Precision Recall (PR)* have also been used to evaluate methods for the task of predicting visual interestingness. *Precision* is the fraction of the items retrieved which are relevant to the user's information need, while *recall* is the fraction of items which are relevant to the query that are successfully retrieved:

$$precision = \frac{\{relevant\ items\} \cap \{retrieved\ items\}}{\{retrieved\ items\}} \tag{5}$$

$$recall = \frac{\{relevant\ items\} \cap \{retrieved\ items\}}{\{relevant\ items\}} \tag{6}$$

The values of precision and recall ranges from 0 to 1. *Average precision* combines recall and precision for ranked retrieval results. Specifically, it is the precision averaged across all values of recall between 0 and 1.

Top_N *score* — The Top_N score [Gygli et al. 2013] was proposed as a measure to quantify how well the top N images automatically selected agree with the human ranking. It is formulated as:

$$Top_N = \frac{\sum_{k \in P_N} s_k^*}{\sum_{k \in H_N} s_k^*} \tag{7}$$

where s_k^* is the ground truth interestingness score of image k, P_N is the set of N images predicted as more interesting and H_N is the set of N images judged as more interesting by human annotators.

Accuracy — Accuracy is a common evaluation measure for assessing the performance of a classifier. It is defined as the overall percentage of correct estimations:

$$Accuracy = \frac{\sum_{k \in P_N} s_k^*}{\sum_{k \in H_N} s_k^*} \tag{8}$$

In the context of subjective labels as interestingness, this measure can be used when annotations are collected in a pairwise way. In this case, accuracy indicates the percentage of item pairs where the most or least interesting item over the two was correctly predicted.

5. COMPUTATIONAL UNDERSTANDING OF VISUAL INTERESTINGNESS AND COVARIATES

Naturally, given the subjective and human-centered nature of the concept of interestingness and its covariates, there is a significant corpus of literature from humanities and social sciences dedicated to these topics. One major drawback of these studies is, however, that they tend to be not scalable, as they allow to classify or annotate few media items only. To understand and predict interestingness at scale, consistent efforts from the multimedia and computer vision communities have been made in recent years to replicate the human judgements in automated, machine-based systems. In this section we review the most relevant approaches.

5.1. Interestingness

In the recent years, computer science researchers have been trying to model interestingness as the property of multimedia data to arise the interest of the observer. In general, interestingness models are based on (1) a training set of media items annotated with some sort of "interestingness" degree, (2) audio/visual features designed to expose interesting patterns in the training data, generally modeling interest "factors" such as novelty and (3) machine learning algorithms employed to learn models able to distinguish between interesting and not-interesting items. While the main relevant datasets on interestingness have been reported in Section 4.2, we present below some of the main approaches in terms of features and algorithms for estimating visual interestingness. Additionally, we report the main approaches proposed to directly estimate concepts related to interestingness, such as emotions or memorability. A summary of the main methods and articles is to be found in Table VIII.

Table VIII. Computational understanding of visual interestingness: a summarizing view of the main methods and articles. The method description column will also present supervised (S) or unsupervised (U) approaches, statistical (T) approaches or papers that survey multiple methods (M).

Citation	*Method description*	*Datasets*	*Metrics*
[Gygli et al. 2013]	(S) Proposes features based on novelty/unusualness, aesthetic value and general user preference and tests according to their context: strong, weak and no context.	Multiple datasets	Average precision, Spearman rank and Top5
[Soleymani 2015]	(S) Uses HOG, LBP and GIST as features, and a regression with sparse approximation is implemented.	visInterest	Pearson's ρ and RMSE
[Jiang et al. 2014]	(S) Employs visual (HSV histograms, GIST, SIFT etc.), audio and high-level features (based on Classemes, ObjectBank and style attributes).	Youtube dataset	Prediction accuracy
[Gygli and Soleymani 2016]	(S) Employs simple visual features, GIF features, Google Cloud Vision API created features, Sentiment features and CNNs with spatio-temporal features (C3D).	gifInterest	Spearman's rank and RMSE
[Grabner et al. 2013]	(S) Employs low-level features such as GIST, HOG etc., emotional, complexity and novelty attributes and a human consensus based learning system.	Webcam dataset	Top3 and Average precision
[Liu et al. 2009]	(U) Proposes an unsupervised approach that compares video frames with photos in Flickr collections, based on SIFT keypoint features for matching, interestingness being calculated as the degree of similarity between individual frames and the Flickr collections.	n/a	User study
[Demarty et al. 2017a]	(M) Presents an overview of the methods in the MediaEval 2016 Predicting Media Interestingness task [Demarty et al. 2016].	MediaEval2016	MAP
[Fan et al. 2016]	(T) Proposes a statistical model of human perception by joining several datasets with partially overlapping perceptual concepts.	Multiple datasets	Spearman's rank and AUC
[Almeida et al. 2017]	(S) Proposes four learning-to-rank methods (Ranking SVM, RankNet, Rank-Boost, ListNet) applied on motion and audio features with a Borda fusion scheme.	MediaEval2016	MAP
[Parekh et al. 2018]	(S) Proposes a pairwise comparison algorithm for emulating human annotations coped with AlexNet fc7 deep features.	MediaEval2017	MAP
[Liu et al. 2018]	(S) Employs feature encoding and feature contribution analysis of low-level content descriptors.	MediaEval2017	Weighted RMSE and MAE

Predicting Interestingness of Images — [Gygli et al. 2013] studied interestingness of images on two crowdsourced datasets (see Table V), and analyzed the impact of three different factors on visual interestingness: *unusualness/novelty*, *aesthetics* and general preference of scene type. The three factors were modeled as follows. A Novelty feature

was designed through Local Outlier Factor. The aesthetic value was approximated using values for colorfulness [Datta et al. 2006], arousal [Machajdik and Hanbury 2010], complexity (based on JPEG size), contrast and edge distribution [Ke et al. 2006]. General preferences were calculated using a RBF-kernel Support Vector Regressor on raw RGB pixel values, GIST, spatial pyramids on SIFT histograms and color histograms. Both datasets were dominated by the general preference approach. Overall, they report a performance of Spearman's ρ=0.71, AP=0.83, Top_5=0.68 on the scene category dataset and Spearman's ρ=0.60, AP=0.73 and Top_5=0.82 on the generic dataset. [Soleymani 2015] evaluated the performance of a general interest predictor on the collected dataset (see Table V) using features proposed in [Khosla et al. 2012], namely HOG, LBP and GIST [Oliva and Torralba 2001] and a regression with sparse approximation of data [Noorzad and Sturm 2012]. A performance of Pearson's ρ=0.44 and $RMSE$=0.13 is reported, where $RMSE \in$ [0,1] (the lower the better). [Fan et al. 2016] present "a paradigm for building generalized and expandable models of human image perception". They show that the use of all data from the fused datasets leads to achieve better performance in terms of using exclusively observed data. They report a performance of Spearman's rank correlation ρ=0.31 and AUC=0.71 on the task of predicting visual interestingness on the fused data. [Marquant et al. 2018] proposed three methods for assessing the interestingness of image regions and apply these methods to the immersive content domain. The methods were based either on traditional, low-level approach or on semantic content detection algorithm, based on fast-RCNN [Girshick 2015].

Predicting Interestingness of Videos — [Jiang et al. 2013] used visual, audio and high-level attribute features to predict interestingness of videos on the collected YouTube dataset (see Table VI). Visual features included Color Histogram in HSV, SIFT, GIST, etc., while attribute features included Classemes [Torresani et al. 2010], ObjectBank [Li et al. 2010], and Style attribute [Murray et al. 2012]. They show that the fusion of audio and visual features obtained through a multi-modal fusion and given as an input for a SVM-Ranking algorithm, greatly improves the performances of interesting video retrieval. The obtained performance on the YouTube dataset, measured in terms of accuracy, is 71.4%. [Gygli and Soleymani 2016] used several visual features to predict GIF interestingness on the collected gifInterest dataset (see Table VI). Models trained with *Sentiment* features [Jou et al. 2015] show the highest predictive performance compared to those trained on a C3D [Tran et al. 2015] - i.e., CNN features with spatio-temporal convolutions trained for action recognition in videos - and those trained on simple visual features such as balance, brightness, etc. The performances are reported with Spearman's rank correlation ρ metric and RMSE, where the best results achieved with their method are ρ=0.53 and $RMSE$=0.17.

Predicting Interestingness within a Video — One of the first works using a crowdsourced dataset for interestingness modeling in *video* streams is the one from [Grabner et al. 2013]. The authors built a model to detect interesting moments in surveillance video streams based on the image Gist feature, and additionally computed low-level, emotion, complexity and novelty features. Complexity and novelty were calculated using respectively the file size after compression and the Local Outlier Factor [Breunig et al. 2000]. Their computational analysis showed that novelty/abnormality of visual events was one of the main factors related to interestingness, thus confirming the findings of previous studies in experimental psychology [Berlyne 1960a; Chen et al. 2001; Silvia 2005]. However, the authors found that novelty alone would not completely explain visual interestingness. The authors called for the necessity of a semantic-aware interestingness model to reach a complete understanding of visual interestingness: for example, simple abnormal events such as cloud formations were not considered interesting, while usual events such as the raising of the Tower Bridge were labeled as

interesting. The best performance obtained by [Grabner et al. 2013] on the collected Webcam dataset (see Table VII) are Top_3 median: 0.72 and AP median: 0.36. In a follow up work, [Gygli et al. 2013] achieved comparable performance on the same dataset, specifically Top3 median: 0.66 and AP median: 0.39. To overcome the training annotation issue, a completely unsupervised approach for interesting video frame detection was proposed by [Liu et al. 2009]. In this work, SIFT features were used to match a video frame to photos in Flickr photo collections. The amount of matching Flickr images was used as estimator of interestingness, following the intuition that Flickr users tend to curate their photo collections to make them interesting [Krages 2012]. The interestingness of each frame was then calculated as the number of features in common with the photo collection, weighted by the similarity of the position of the SIFT feature in the photo and frame. The approach was evaluated through a user study where participants (5 graduate students were selected) were asked to watch each video and rate the corresponding generated interestingness curve. A score between 1 and 5 could be given (the higher the better). All videos received a score equal or greater than 4.

So far, all the articles presented in this section have used different sets of images or videos and evaluation methods for their performance, thus making it difficult to compare the different algorithms. The first initiative to provide a common evaluation framework is the MediaEval 2016 Benchmarking Initiative for Multimedia Evaluation with the Predicting Media Interestingness Task [Demarty et al. 2016] (see Section 4.2). The 12 participating teams addressed many different methods for classification and description. The descriptors included low level features like color histograms, GIST, Histogram of Oriented Gradients, motion histograms [Constantin et al. 2016], deep features like the output of the AlexNet fc7 and prob layers and VGG [Shen et al. 2016], high-level attributes such as the presence and predominance of faces, style attributes and emotion-based attributes [Liem 2016] and many fusion approaches. As for the classification algorithms, different variations of SVM were widely used [Shen et al. 2016], some in late fusion implementations [Lam et al. 2016], while some teams decided to apply deep learning techniques for classification [Xu et al. 2016]. The best performing approach for the image subtask was recorded by [Liem 2016] and [Shen et al. 2016], with a final MAP score of 0.2336. The first team used a rule-based approach with simple human face information features (color and sizes) concluding that faces attract attention and interest, while the second one used SVM training with RBF kernel on features extracted from the fc7 layer of AlexNet. The best MAP score, 0.1815, for the video subtask, was recorded by [Almeida 2016], who used Histograms of motion patterns in a normalized confidence scoring system. For an in-depth analysis of approaches, the reader may refer to [Demarty et al. 2017a]. We can also identify in literature some relevant approaches which were validated on these data. For instance, [Almeida et al. 2017] employed Ranking SVM, RankNet, Rank-Boost and ListNet learning-to-rank on motion and audio features. The authors obtained a MAP of 0.1997 via a Borda fusion of their best performing single-modal systems.

The continuation of this task was the MediaEval 2017 Predicting Media Interestingness Task [Demarty et al. 2017b], which provided more data to the participants. Again, some teams used features such as color, histograms, GIST, HOG, LBP [Yoon 2017], while a high degree of interest was shown for features from domains that were found to be connected with interestingness, like aesthetic and style features, affective impact, visual attention, saliency maps and novelty. Context information and textual metadata from movie descriptions were also used, e.g. in [Berson et al. 2017], as well as training approaches with pair-wise comparison between sequences from the same trailer [Parekh et al. 2017]. In terms of classification, teams used certain variations of SVM learning systems [Ben-Ahmed et al. 2017], logistic regressions [Permadi et al. 2017], learning to rank algorithms [Almeida and Savii 2017] or deep ranking mod-

els [Wang et al. 2017]. Late fusion was used for boosting performance [Almeida and Savii 2017; Constantin et al. 2017]. The best performing system for the image subtask achieved a MAP@10 score of 0.1385 and a MAP score of 0.3075 and consisted of a logistic regression algorithm performed on LBP, HOG and AlexNet fc7 features. For the video subtask [Ben-Ahmed et al. 2017] had the best scores, MAP@10 = 0.0827 and MAP = 0.2094, with a sigmoid kernel SVM and with deep visual and audio features (VGG and SoundNet). As examples of relevant approaches from the literature validated on these data we can mention the work of [Parekh et al. 2018] which emulates the human processing of data annotations via a pairwise comparison of AlexNet fc7 layer features, or [Liu et al. 2018], who applied a jointly optimized feature selection algorithm for predicting interestingness and emotions. Another work that uses this dataset is [Ben-Ahmed and Huet 2018], where the authors propose a method that uses deep multimodal features, extracted from ResNet-125, LSTM and Soundnet. This methods recorded a final MAP score of 0.2122, above the best performance achieved in the MediaEval competition.

5.2. Affective Value and Emotions

A large body of literature has focused on affective analysis of visual content. The majority of work has focused on automatically predicting users' emotions from facial expression, such as from webcam videos, thus focusing on the emotions expressed by a user. Yet, a relatively small set of works focused on predicting emotions evoked from visual content, which is subjective and it may vary according to the subject observing the content. In this survey, we are interested in this latter group of works. Emotions can be considered either in a dimensional or categorical representation. The dimensional representation allows to potentially represent any possible emotion due to a continuous mapping. Still, emotions themselves are not quantitative variations on abstract dimensions, there is a limited number of distinct emotions [Silvia 2006].

Interestingness and Arousal — [Gygli et al. 2013] used arousal among several other features to predict visual interestingness, where the evoked arousal from an image is computed by using pixel brightness and saturation [Valdez and Mehrabian 1994], i.e. $Arousal = \sum_p 0.31 \cdot brightness(p) + 0.60 \cdot saturation(p)$. This feature showed a relatively high prediction power in the case of weak context with images from scene categories, while almost no prediction power in the context of generic images. The Pearson correlation ρ between the image interestingness and the predicted arousal is equal to 0.43 and -0.02 respectively for the two cases.

Interestingness and Sentiment Features — [Gygli and Soleymani 2016] used several visual features to predict animated GIFs interestingness. They show that Sentiment features [Jou et al. 2015] allow to achieve higher predictive performance (Spearman's correlation $\rho = 0.52$), compared to other visual features such as C3D [Tran et al. 2015] ($\rho = 0.48$) and simple visual features such as balance, brightness, etc. ($\rho = 0.39$).

Interestingness and Categorical Emotion Space — [Gygli and Soleymani 2016] considered various visual features, included *"violence likelihood"* and the degree of the emotions *"sorrow"*, *"anger"*, *"surprise"* and *"joy"*, for the task of predicting animated GIFs interestingness. Still, performance results on this task were reported for the joint set of these features, called Google API, which also include spoof likelihood, presence of faces, etc. Therefore, it is not possible to draw any conclusion based purely on the link between interestingness and violence, sorrow, anger, surprise or joy from this study.

Predicting Emotions — A huge body of literature has focused on predicting emotions evoked by visual content, where a large effort has been put to identify effective features for bridging the semantic gap in the affective estimation. *Predicting Emotions in a Dimensional Emotion Space*: [Sartori et al. 2015a] investigated the effect of color

combination on the emotional message conveyed by abstract paintings. [Acar et al. 2015] showed that using CNN based mid-level features and temporal features, fused at decision level, allows to achieve better performance on the task of affective video estimation in the VA space with respect to using low level audio-visual cues. *Predicting Emotions in a Categorical Emotion Space*: [Zhao et al. 2014] proposed a novel set of features based on high level artistic principles, like *harmony* or *proportion* which show a higher predictive power and which are more interpretable by humans with respect to artistic elements. CNN approaches have also been used for affective estimation [Peng et al. 2015]. [Jou et al. 2014] propose a multi-task approach to jointly estimate the scores for the discrete set of 17 emotions, thus exploiting the inherent correlation between labels, like for example "contentment" and "amusement". Some papers have analyzed the performance of their systems with respect to both categorical and dimensional emotional spaces. For example, [Mo et al. 2018] creates the HHTC feature set for recognizing 6 base emotions and classification along the arousal-valance dimensions.

Detecting Boredom in Video — During the Mediaeval 2010 Affect Task a few works have attempted to address the automatic estimation of boredom from videos [Soleymani 2010]. In this challenge, boredom is related to viewer's sense of keeping attention and sense of passage of time. Boredom is also deemed as an indicator that the video quality perception is low. The task consists in identifying video content which causes the viewer to feel bored rather than engaged. Some of the approaches in this domain included low-level features like key lighting and color variance [Soleymani 2010] or high-level concepts like cuteness, dynamism, humor, interactivity and popularity [Shi and Larson 2010], the latter showing that humor is the attribute which most helps in predicting the level of viewer engagement.

5.3. Aesthetic Value

The computational assessment of visual aesthetic quality is gaining importance in the digital age. There is a growing need to classify high volumes of data and provide the users with a better user experience by showing more aesthetically pleasing results. Thus, several research directions have emerged, some dealing with the perceived aesthetic value from a psychological point of view (see Section 3.3), others, in the field of photography, with photographic rules such as composition, depth of field, color theories and themes etc. or using low-level or high-level computer vision concepts [Ke et al. 2006].

Predicting Aesthetic Value — Pioneers in this field were [Datta et al. 2008]. Using datasets collected from photo annotation contest websites such as DPChallenge.com or Photo.net, these works designed low-level features inspired by photographic rules (e.g. depth of field) and effectively predict photo aesthetic values. Some of these photographic rules were exploited further, such as in [Hernández-García et al. 2016]. Recently, researchers have shown that convolutional neural networks [Mai et al. 2016] can effectively be used for visual aesthetic value prediction, showing promising improvements.

Interestingness and Aesthetic Value — Aesthetics was studied and proven to be correlated with interestingness from a computer vision perspective in many papers like [Dhar et al. 2011; Gygli et al. 2013; Hsieh et al. 2014]. The authors of [Dhar et al. 2011] extracted three types of high-level attributes to determine image interestingness and aesthetic quality - "compositional", "content" and "sky-illumination" attributes. The "compositional" attributes taken into account are generally closely related with the field of image aesthetics - rule of the thirds, low depth of field, opposing color theory as described in [Schloss et al. 2013] and the presence of salient objects [Liu et al. 2011]. The "image content" attributes used in this study were: presence of people and

portrait depiction (the presence of a face and whether the main subject is a single large face or not) and presence of animals, indoor-outdoor classification and objects scene classification, all of these last three concepts being computed with SVM classifiers. As for "sky-illumination", the focus lay on three basic categories: clear, cloudy and sunset skies. For both aesthetics and interestingness the resulting classifiers performed very well, and the best was a SVM system trained using both high and low-level descriptors inspired by [Ke et al. 2006], while the high-level attributes outperformed the low-level ones. Several factors were studied by [Gygli et al. 2013] in correlation with interestingness, and aesthetics was the second best correlated factor, after perceived memorability, with a Spearman rank correlation of 0.59 when comparing the ground truth data. The extracted features for detecting aesthetics were: colorfulness as presented in [Datta et al. 2006], arousal [Machajdik and Hanbury 2010], complexity as a measure of JPEG compression size, contrast and edge distribution [Ke et al. 2006]. Flickr interestingness score was used in the context of video aesthetic modeling. Using Flickr's interestingness annotations and computational aesthetics techniques, [Redi and Merialdo 2012] trained a model for image social interestingness detection to retrieve the best shots from the NHK dataset [Takahashi and Sano 2013] (see Table VI).

5.4. Memorability

"Today's expansion of infographics is certainly related to one of the everyday life idiom A picture is worth a thousand words and to the need of providing the fastest possible knowledge transfer in the current information overload age" [Siarohin et al. 2017].

Predicting memorability — By an extensive study based on crowdsourced data, [Isola et al. 2011b] showed that "memorability is an intrinsic property of an image" that can be automatically predicted. [Khosla et al. 2015] show that deep features help to achieve a correlation ranked in predicting memorability closed to the correlation rank between memorability scores obtained by using two independent sets of viewers for the same image. In general, it was shown that high level attributes allow to achieve a higher accuracy in predicting memorability with respect to low level cues based on color or simple image features, like mean hue [Isola et al. 2014] or mean brightness [Dubey et al. 2015]. However, some exceptions do exists. [Shekhar et al. 2017] show that color features could obtain superior results compared to semantic and deep features. Memorability has also been used, e.g., in [Fei et al. 2018] via a fine-tuned AlexNet model, as a feature for creating video summaries.

5.5. Novelty

The problem of detecting novelty or related topics like unusualness has been extensively studied in image and video analysis with the final goal to identify unusual spatio-temporal patterns. These include anomalous behaviors or suspicious activities from traffic surveillance videos, detection of unusual regions on images for medical purposes, public space monitoring or home activities [Varadarajan and Odobez 2009; Zhao et al. 2011].

Predicting novelty — A high number of approaches were based on clustering and assigning a membership threshold value for each cluster created, therefore, signaling out the media samples that were not cluster members as novel [Yong et al. 2013].

Interestingness and Novelty — [Gygli et al. 2013] propose two methods to compute unusualness/novelty of an image, based on global and partial descriptors. An important observation from this study is that the proposed measures of unusualness can mostly capture information about the variance of the data, but it cannot provide information about the a priori knowledge of the observer, which is what supposedly has an influence on visual interestingness. The Local Outlier Factor (LOF) [Breunig et al. 2000] algorithm was successfully used by [Datta et al. 2006] to classify the aesthetic

proprieties of an image. Using the same LOF idea, [Gygli et al. 2013] tried to predict interestingness for a set of images, integrating this feature inside an unusualness descriptor. The LOF algorithm used a 10-distance neighborhood and the sub-component features were: raw RGB pixel values, GIST [Oliva and Torralba 2001] and Spatial Pyramids on SIFT Histograms [Lazebnik et al. 2006]. The performance of this descriptor was very good on the strong context dataset, having the highest Spearman's rank correlation (0.29) and Average Precision (0.35) and the second highest $Top5$ score (0.51) among all the descriptors.

5.6. Complexity

Information theory gave an idea of complexity based on the entropy of data and on certain derived proprieties of the compressed data, based on the works of [Huffman 1952; Li and Vitányi 2009]. Image complexity measures have been used in a number of image processing and computer vision related problems like image region segmentation [Rigau et al. 2007], data compression [Huffman 1952], benchmark systems for target recognition difficulty [Peters and Strickland 1990].

Predicting Complexity — Some authors used complexity measures that are object-dependent such as the number of objects or targets in an image [Bhanu 1986], contrast metrics between the subject of a media shot and its background [Peters and Strickland 1990], or edge-dependent measures such as the one implemented by [Yu and Winkler 2013]. Fuzzy models were also employed by [Cardaci et al. 2009]. Numerous approaches used compression or entropy-based measures in works such as [Perkiö and Hyvärinen 2009]. Three measures of visual clutter were used by [Rosenholtz et al. 2007] for measuring image complexity and determining how easy it would be for a human observer to retrieve objects from a scene.

Interestingness and Complexity — In interestingness prediction, complexity was often determined as a function of compressed image size, and studied in connection with interestingness in the works of [Grabner et al. 2013; Gygli et al. 2013]. In [Gygli et al. 2013] complexity was calculated as a ratio between the JPEG compressed image size and the original uncompressed image size, while [Grabner et al. 2013] used the file size of the frames in the image sequences, compressed as PNG files and all these studies shown a good performance for the complexity feature. The opposite concept to complexity, simplicity, was found by [Ke et al. 2006] to be one of the most important attributes of high-quality images, stating that it is essential for a professional to clearly separate the background from the subject of the picture, by different techniques such as short focus, color contrast or lighting contrast. Based on these observations several computer vision techniques have been employed in order to obtain a simplicity measure. [Ke et al. 2006] used spatial distribution of edges, concluding that amateur photos have a cluttered background and that professional images look more colorful and have a lower unique hue count, because of the non-intrusive background. [Luo and Tang 2008] used a background color distribution of the RGB channels, obtaining separate values for amateur and professional shots, while [Yeh et al. 2010] calculated the ratio of the Regions of Interest (ROI) segments compared to the whole image area as a complexity measure.

5.7. Coping Potential

Reducing the cognitive load is crucial when transmitting rich information, and understanding the nature of coping potential and related concepts is fundamental to quantify the extent to which a multimedia item can be easily processed and cognitively digested. However, predicting and modeling coping potential is a non-trivial task.

Predicting Coping Potential — Apart from [Soleymani 2015], who tried to predict coping potential from visual features (with the following accuracy values $r^2 = 0.06$, $\rho =$

0.27 and $RMSE = 0.11$), no computational studies have been made regarding coping potential, and in general no concept from this group has been studied from a pure computational perspective. This lack of studies can be attributed to the high degree of subjectivity of these concepts and perhaps to the high challenge of transforming the definitions and the conclusions of related psychological studies into computational algorithms.

5.8. Visual Composition and Stylistic Attributes

Visual composition plays an important role in aesthetic perception [Ke et al. 2006] and in generating interest [Jiang et al. 2013]. In literature, compositional attributes such as color distribution or texture distribution have been implemented as features to predict aesthetic value [Datta et al. 2008], memorability [Isola et al. 2011b], popularity [Khosla et al. 2014], creativity [Redi et al. 2014], and interestingness [Jiang et al. 2013]. In this section, we will limit our analysis of compositional attributes to their relation to interestingness prediction.

Interestingness and Realism — Realism was considered one of the three important factors in differentiating high-quality, professional photos from amateur ones by [Ke et al. 2006]. The tendency among the studied image sets was that professional photographs had a surreal look, while amateur ones had a realistic look; the surreal impression is considered as a result of the use of special color palette or balance, camera and shot settings or subject composition techniques. Human interest can also be aroused by abstract works of art, poems and other concepts, as long as they are deemed as understandable, as pointed out in [Silvia 2008]. So far, no definite computer vision study has been made regarding the interest of real or abstract multimedia data.

Predicting Naturalness — Naturalness of images can be either computed as a linear combination of saturation values for skin, sky, and grass pixels [Huang et al. 2006], or predicted from crowdsourced judgments, as in [Soleymani 2015] ($r^2 = 0.32$ $\rho = 0.57$ and $RMSE = 0.18$).

Interestingness and Photographic Composition — The photographic composition was considered as a possible cue for interestingness in [Jiang et al. 2013], the authors defining "photographic style attributes" as a high-level descriptor for their video interestingness prediction system. The descriptor was formed by concatenating outputs from the 14 compositional attributes described in [Murray et al. 2012], a paper dealing with aesthetic visual analysis, and predicted by several features and their combinations like Color histograms, SIFT and LBP, but the final conclusion was that the accuracy for these features is low. In a work that approaches both videos and images, in correlation with aesthetics, [Wang et al. 2013] have indeed shown that there is a big difference for the style descriptor when predicting Spearman's rank correlation for images (0.23) vs video prediction (-0.01). Some photographic attributes were used to predict image interestingness in other papers [Dhar et al. 2011; Redi and Merialdo 2012; Liu et al. 2009; Halonen et al. 2011], like Rule of Thirds, Depth of Field, Complementary Colors or Negative Colors, usually with better results. Another approach on photographic composition was taken by [Yoon and Pavlovic 2014], who took 114 features inspired by the works on art theory and psychological studies of [Machajdik and Hanbury 2010]. The authors suspected that in certain professionally edited videos such attributes might help predict interestingness better than in the case described by [Jiang et al. 2013]. The style features performed well on the DEAP [Koelstra et al. 2012] database when they are used as isolated features, although they were not the best performing features under analysis. In literature [Machajdik and Hanbury 2010; Schifanella et al. 2015], texture information is generally extracted through Local Binary Patterns [Ojala et al. 2002], or as Haralick's features [Haralick 1979]: "Entropy", "Energy", "Homogeneity", "Contrast of the Gray-Level Co-occurrence Matrices". [Hsieh

et al. 2014] found that "texture and color features have best and worst performance, respectively", on predicting visual interestingness. In particular, it was found that "colors such as red and violet can evoke greater arousal" [Hsieh et al. 2014], which plays an important role in re-sharing behaviors. Individual color distributions can be automatically extracted from images through color names, i.e. frequency distributions of pixels over pre-defined color clusters [Van de Weijer et al. 2007], or color moments [Yu et al. 2002], i.e. the values of the first, second and third moments of the pixel distributions in the red, green and blue channel. [Datta et al. 2006] propose to measure the global colorfulness of an image as the Earth Mover's distance [Rubner et al. 1998] (in the LUV color space) of the color histogram of the image itself to a uniform color histogram. This measure was also used in [Gygli et al. 2013] and proved to be positively related with interestingness.

5.9. Social interestingness

In the recent years we have assisted to a rising importance of social media with respect to mainstream media. A recent study from Pew Research Center [Center 2016] reports that the majority of U.S. adults get news on social media. [Redi et al. 2015] showed that beautiful items can remain undiscovered due to the lack of visibility in a photo sharing platform. In this context, it is of uttermost importance for users and companies to understand the underlying mechanisms of social media, in order to maximize the chances for an item to be discovered, liked, reshared.

Predicting Virality — Virality is usually computed based on users' reactions to posts, though literature does not offer a universal formulation for computing virality. [Guerini et al. 2013] investigated the contribution of image low-level cues on virality. [Deza and Parikh 2015] show that automatic systems perform better than humans at predicting virality, reporting the importance of using high level-features rather than low-level ones for achieving a better performance. [Alameda-Pineda et al. 2017] proposed a deep learning architecture to localize the area in an image which is mostly responsible for resharing.

Predicting Popularity — [Dhar et al. 2011] investigated three types of factors to determine image social interestingness, where the Flickr metric [Butterfield et al. 2014] is used as score, namely: compositional attributes (modeled using computational aesthetics techniques), content (e.g. people or salient objects) and sky-illumination type, such as clear, cloudy and sunset. The reported plots of Average Precision show that these high level describable attributes perform better than low level features at the task of social interestingness prediction. Some other papers use deep neural network features for predicting popularity, such as [Fontanini et al. 2016] (AlexNet and DeepSentiBank) and [Chen et al. 2018] (VGG-16, VGG19, ResNet-152, Inception-X and NasNet).

Interestingness and Popularity — [Ung 2011] investigates the link between popularity, interestingness and social influence, aiming to estimate the average interestingness that an independent user would express for a specific online content (e.g. a story), in the absence of the social influence. They show that the effect of *social influence* and *visibility* tends to favor less interesting stories. Still, to the best of our knowledge, no works have tried to prove this negative link between popularity and interestingness with a computational approach for the case of visual content, like images or videos.

5.10. Creativity

Being able to evaluate the creative value of digital artifacts is extremely important in building effective creative agents and generative art frameworks [Jordanous 2012]. Moreover, visual creativity detection has wide application in the context of multimedia retrieval and recommendation. In some contexts, image/video search engines might

want to rank multimedia items not only according to their relevance to a query, but also according to their creative value; image/video sharing platforms such as Instagram or Vine could recommend specific multimedia items according to their intrinsic creative value, as opposed to their popularity in the platform. Intuitively, creativity and interestingness should be intrinsically linked, however, the correlation between these two concepts has not been studied using a computational approach.

Predicting Creativity — Despite its importance, creativity has been rarely explored in computer vision literature. Based on the definitions presented in Section 3.10, [Redi et al. 2014] created a method to determine creativity in micro-videos (videos that last 6 seconds or less). Seven groups of features were extracted for videos, based on the following general types: "scene content", "compositional/photographic techniques", "filmmaking technique", "visual emotional affect", "audio emotional affect", "visual novelty" and "audio novelty". Some important interestingness factors, such as symmetry, low depth, skin color, create a clear separation between interestingness and creativity.

5.11. Humor

Humor plays an important role in human social interactions, communications and has a positive effect on health, happiness and overall human lifestyle. Therefore, it is important for computer vision systems to predict humorous media shots, with direct applications in certain areas such as recommendation systems that could return humorous images or videos with a higher priority and to create social tools that help users create or select the appropriate posts.

Predicting Humor — There have been many approaches regarding the prediction of humor in text-based systems [Davidov et al. 2010]. Few approaches, however, have dealt with visual humor and computer vision algorithms for the detection of humorous images or videos. [Chandrasekaran et al. 2016] created a system that both predicts humor and can alter how funny a scene is. Instance-level descriptors that predict the instances where object categories are more likely to occur and scene-level features that deal with object descriptors across the scene were extracted from the images, and the results show that the trained SVR system performs above the baseline for predicting humorous images.

Predicting Irony and Sarcasm — The study of irony has been concentrated on textual information and used in special cases such as social media detection and tweets [Riloff et al. 2013], audio data with several approaches to intonation and pitch [Shapely 1987] and facial gestures, as identified by [Attardo et al. 2003], such as blank face, eyebrow movement, eye movement, winking, nodding and smiling. No approach has yet been made regarding a connection between irony and interestingness. Sarcasm, a subcomponent or special type of irony, as presented by [Kreuz and Glucksberg 1989], has usually been studied as a synonym for irony [Attardo et al. 2003].

Humor and Interestingness — [Jiang et al. 2013] noticed that video sequences labeled as humorous had a better interestingness score overall. However, no attempt was made to determine this propriety through a computer vision approach, the authors noting that, provided some high-level features were developed, the final results of the prediction algorithm could be improved. [Shi and Larson 2010] used several high-level concepts like humor and cuteness in a paper dealing with boredom detection. Humor, determined by laughter detection, as presented in [Stolcke et al. 2008] was the best performing feature, scoring $Kendall - \tau = 0.19$ and $p - value = 0.01$.

5.12. Urban perception

Understanding how people perceive the urban space is crucial for the future of urban design and the effectiveness of smart cities. The practice of collecting human perception of urban places has commonly been used in social and urban planning stud-

ies [Lynch 1960]. Still, it was conducted through manual surveys and it was naturally limited to the number of participants to the study. Although some existing works use computer vision to predict urban perception dimensions, to the best of our knowledge, no works. have been specifically investigating the link between urban appearance and the visual or social interestingness of a geographic area.

5.13. Saliency

Predicting Saliency and Attention — Over the last decades, a large body of literature focused on modeling visual attention, particularly saliency-based attention. For a deeper look at these approaches, we refer the reader to [Borji and Itti 2013; Borji et al. 2015]. Giving an extensive overview of these works is out of the scope of this survey. The role of attention, often measured by recording eye movements [Hoffman and Subramaniam 1995], was studied in relation to interestingness covariates such as memorability [Mancas and Le Meur 2013] and emotional valence [Hamed et al. 2015]. Approaches predicting visual attention using a computational approach often use saliency and visual interestingness as interchangeable terms, see [Chaabouni et al. 2017] for example. [Bylinskii et al. 2015a] measure the effectiveness of different types of visualization by aggregating eye movement metrics, while [Le Meur et al. 2011] show that aggregated eye fixations measures can help in personalized image ranking.

Saliency, Attention and Interestingness — In general, only few seminal works investigated the role of saliency and attention in visual interestingness: [Onat et al. 2014] studied localized interestingness, attention and saliency, showing that there is a high inter-relation between these 2 quantities; [Bylinskii et al. 2016] also explored the role of text interestingness in scenes to predict eye fixations.

6. APPLICATIONS

People who work in entertainment, art or education must know how to arouse a feeling of interest or to evoke certain emotions when creating content for a specific audience. This property is also desirable for intelligent multimedia systems. The recent attention brought by scientists in various fields like multimedia, robotics, and artificial intelligence to visual emotions and subjective perception gives the premises to bring technology closer to people desires and needs. Recent studies in psychology and neuroscience reveal the strict connection between the human cognitive and emotional aspects. Cognitive processes give raise to emotions [Ellsworth and Scherer 2003], while emotions have an impact on perception, attention, memory, judgement, and reasoning [Silvia and Warburton 2006]. These premises clearly underline the importance of considering the affective sphere when developing intelligent systems in fields such as education, advertising, e-commerce, entertainment, industry, etc.

In this survey, we have focused on the automatic assessment of visual interestingness, highlighting the connection between interestingness and other concepts, ranging from emotions to aesthetics. While the understanding of visual interestingness is our ultimate goal, our belief is that the comprehension of the link between these concepts will help us progress into other directions, like the automatic assessment of other subjective properties related to interestingness. This section offers an overview of the main applications which may benefit from the advances in the modelling of visual interestingness and, in general, of visual subjective properties.

Media Summarization. For example, the ability to automatically assess the interestingness of the different video moments and to extract the video highlights would have an impact in many industrial applications. Video service providers would be able to provide significant video preview able to attract users' interest. Users would be allowed to perform an efficient video browsing, for example by being able to trim redundant or non-interesting parts of videos. Also, in applications like visual summarization

from lifelog camera video streams, we can easily imagine the different impact of creating summaries considering also the emotional aspect and the aesthetic assessment of the visual content, rather than simply reporting the most recurrent objects and people in a video. A similar concept applies to users photo collection such as holiday pictures. A user may ask the system to retrieve the most interesting or beautiful pictures out of a large picture collection. A significant line of works has been following this trend by proposing the use of generic measures of visual interestingness in order to create video summaries closer to human performance [Gygli et al. 2014, 2015] or focusing on emotion-oriented summarization [Xu et al. 2018].

Media Recommendation. Systems able to surface interesting content for users can be employed in recommendation systems. For example, for photo sharing platforms, where users are not able to consume the huge amount of content in its entirety, being able to suggest interesting items is key. The automatic assessment of subjective concepts like aesthetic allows to surface hidden beautiful images in photo-sharing communities like Flickr, in contrast to methods using metrics based on social interest values like image popularity, as shown in a recent work from [Schifanella et al. 2015]. In the broader case of media recommendation systems, user experience may be further improved by suggesting videos closer to user's personality or mood, as well as more appropriate according to user's age or preferences (e.g., violence scenes, explicit content, etc.). Finally, in applications like patient mood monitoring, an automatic assistant may, for example, suggest a safe path evoking feelings of joy, rather than sadness, during an afternoon walk in a city. [Quercia et al. 2014] investigated how people perceive the environment in order to suggest the walking path which can provide a more positive user experience.

Advertising. Advertising is also an application area which may highly benefit from the growing awareness on the emotional impact of visual content. In the case of online ads, these have been long perceived as irritating or annoying, thus fostering the widespread of ad blocking software. In response to that, theories have arisen defending the importance of advertisement, identifying the problem in the "bad advertisement" rather than "advertisement" itself [6]. Specifically, the lack of empathy which is manifested for example with the (in)capacity to act in line with user's mood has been pointed as one of the main causes for "bad" online advertising[7]. Similarly, recent works demonstrate that empathy is one of the key drivers for a positive answer to online advertising, together with informativeness and creativity [Lee and Hong 2016]. These theories have found practical counterparts, for example, empathy has been defined as the key to success in the recent advertisement campaigns in New York[8]. The automatic assessment of the emotional impact of visual content is surely a fundamental step towards a scenario where online advertising can be a non-invasive and a pleasing part of our online navigation. Moreover, companies may want to automatically assess visual content like a new company logo or a recently planned advertising campaign with respect to various purposes, such as estimating their memorability or their visual attractiveness for different target audience. Ultimately, an automatic system may be able to suggest the proper visual changes in order to make the company logo or advertisement more memorable or attractive for that specific audience.

7. CONCLUSIONS AND FUTURE CHALLENGES

In this article we reviewed the main directions for understanding visual interestingness and analyzed its relation with other concepts. We also summarized the type of

[6] http://blog.imonomy.com/ads-designed-empathy-end-ad-blocker-war
[7] https://goo.gl/oHEBm1
[8] http://www.tronviggroup.com/empathy-in-advertising

correlation between interestingness and the other concepts and singled out the positive, negative and still unexplored links, therefore also gaining an insight into controversial topics and into the possibility for future challenges. Visual interestingness and the related concepts, along with their correlation, were analyzed from three different perspectives: The human factor, dealing with studies from psychology and philosophy, User studies, dealing the subjective human assessment of these concepts and finally Computational models designed for their prediction.

Having reviewed a large corpus of research works focusing on the notion of interestingness, we now identify the "missing bits": namely open questions and areas which have been marginally explored by previous work. With this final overview, we hope to foster future research in the field of interestingness understanding and prediction, providing a common ground of possible research directions for researchers in computer vision, multimedia, social computing, and computational anthropology.

Controversies — While the connections emotions-interestingness, novelty-interestingness and memorability-interestingness have been widely explored and tested, the actual relation between some subjective properties and interestingness remains unclear. *Interestingness and Aesthetic appeal*: A positive or negative correlation? User studies trying to investigate the correlation between aesthetics and interestingness came to different conclusions. This inconsistency might suggest that, given the high correlation between novelty/complexity and interestingness, and given the inverted U-shape[9] relation between novelty/complexity and aesthetic appeal, the curve of relationship between aesthetic appeal and interestingness might be non-linear. A systematic study investigating fine-grained interactions between these two concepts would definitely help addressing this open question. *Interestingness and Pleasantness*: While some works propose that high pleasantness drives interest [Ellsworth and Smith 1988], other models, suggest that "people can be interested in disturbing or unpleasant events" [Turner and Silvia 2006]. To disentangle the relation between pleasantness and interestingness, it might be useful to add contextual information into the model, including user preferences and demographics.

Largely Unexplored Relations — Some of the concepts mentioned in this survey have been rather marginally explored in relation to interestingness. *Interestingness and Coping potential*: Although experimental psychology works have shown that the comprehensibility of a multimedia item is positively related to its interestingness, no systematic work on computational understanding of coping potential has been made. Being able to automatically score images and videos according to their comprehensibility would not only contribute to interestingness prediction and understanding. Automatic prediction of coping potential would allow to retrieve the most "easy to understand" items, thus allowing a number of useful applications in multimedia retrieval for news and education. *Interestingness and Creativity*: The link between creative artifacts and the interestingness they arouse is intuitive – creativity is highly related to novelty, which is, in its turn, a major component of interestingness. However, no previous study has systematically evaluated the relation between creativity and interestingness. Up to which point the creativity of a multimedia item is related to its interestingness? And does coping potential (i.e. the ability to understand the creative artifact) play an important role? These questions still remain unanswered. *Urban Interestingness*: Beyond urban space recognizability and liveliness, little work has been done on understanding what makes urban spaces interesting for individuals. Possible research questions in this direction include: what is urban interestingness, and how

[9] An inverted U-shape pattern on a 2-d plane has low values of Y in correspondence of very low and very high X (very low and very high complexity $\rightarrow$ low interestingness), and high values of Y in correspondence of median values of X (some complexity $\rightarrow$ high interestingness)

does this relate to visual interestingness? What are the urban elements that make a street or neighborhood visually interesting? *Interestingness and Humor*: Humor and interestingness: how are visual properties arising laughter, such as humor, irony or sarcasm related to the notion of visual interestingness? To the best of our knowledge, no existing work has deeply investigated these links. *Subjective Interestingness in Context*: So far, we considered interestingness (and many other subjective concepts) as a "universal" concept: many of the works mentioned assume that the subjective perception of a multimedia item is equal for every observer. However, previous studies in experimental psychology demonstrated that demographics, the education, the experience of a user change the way in which we assign subjective judgments to works of art. We hope in the future to see computational methods studying and predicting interestingness in relation to user characteristics: this would have the double advantage of (1) producing interesting algorithmic insights regarding how cultural/sociological aspects impact our subjective perception, thus contributing to the field of social and cross-cultural psychology and (2) making multimedia interestingness prediction and retrieval frameworks more accurate and personalized. *Cultural Aspect*: The question on how cultural aspect influences interestingness is related to the non universality of the feeling of interest. We did not find studies in literature which properly address this question.

REFERENCES

Esra Acar, Frank Hopfgartner, and Sahin Albayrak. 2015. Fusion of learned multi-modal representations and dense trajectories for emotional analysis in videos. In *IEEE International Workshop on Content-Based Multimedia Indexing*. IEEE, 1–6.

Peter P Aitken. 1974. Judgments of pleasingness and interestingness as functions of visual complexity. *Journal of Experimental Psychology* 103, 2 (1974), 240–244.

Xavier Alameda-Pineda, Andrea Pilzer, Dan Xu, Nicu Sebe, and Elisa Ricci. 2017. Viraliency: Pooling Local Virality. In *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 484–492.

Jurandy Almeida. 2016. UNIFESP at MediaEval 2016: Predicting Media Interestingness Task. In *MediaEval 2016 Workshop*.

Jurandy Almeida and Ricardo M. Savii. 2017. GIBIS at MediaEval 2017: Predicting Media Interestingness Task. In *MediaEval 2017 Workshop*.

Jurandy Almeida, Lucas P Valem, and Daniel CG Pedronette. 2017. A Rank Aggregation Framework for Video Interestingness Prediction. In *International Conference on Image Analysis and Processing*. Springer, 3–14.

Xesca Amengual, Anna Bosch, and Josep Lluís de la Rosa. 2017. How to Measure Memorability and Social Interestingness of Images: A Review. *International Journal of Pattern Recognition and Artificial Intelligence* 31, 02 (2017), 1754004.

Philip W Anderson and others. 1972. More is different. *Science* 177, 4047 (1972), 393–396.

Stegen R Asher and Kristina L McDonald. 2009. The behavioral basis of acceptance, rejection, and perceived popularity. *Handbook of peer interactions, relationships, and groups* (2009), 232–248.

Salvatore Attardo, Jodi Eisterhold, Jennifer Hay, and Isabella Poggi. 2003. Multimodal markers of irony and sarcasm. *Humor* 16, 2 (2003), 243–260.

Salvatore Attardo and Victor Raskin. 1991. Script theory revis (it) ed: Joke similarity and joke representation model. *Humor-International Journal of Humor Research* 4, 3-4 (1991), 293–348.

Frank Barron and David M Harrington. 1981. Creativity, intelligence, and personality. *Annual review of psychology* 32, 1 (1981), 439–476.

Yoann Baveye, Emmanuel Dellandrea, Christel Chamaret, and Liming Chen. 2015. LIRIS-ACCEDE: A video database for affective content analysis. *IEEE Transactions on Affective Computing* 6, 1 (2015), 43–55.

Olfa Ben-Ahmed and Benoit Huet. 2018. Deep Multimodal Features for Movie Genre and Interestingness Prediction. In *2018 International Conference on Content-Based Multimedia Indexing (CBMI)*. IEEE, 1–6.

Olfa Ben-Ahmed, Jonas Wacker, Alessandro Gaballo, and Benoit Huet. 2017. EURECOM @MediaEval 2017: Media Genre Inference for Predicting Media Interestingness. In *MediaEval 2017 Workshop*.

Daniel E Berlyne. 1949. Interest as a psychological concept. *British journal of psychology. General section* 39, 4 (1949), 184–195.

Daniel E Berlyne. 1960a. Conflict, arousal, and curiosity. (1960).

Daniel E Berlyne. 1960b. Novelty, Uncertainty, Conflict, Complexity. *Conflict, arousal, and curiosity* (1960),

18–44.
Daniel E Berlyne. 1963. Complexity and incongruity variables as determinants of exploratory choice and evaluative ratings. *Canadian Journal of Experimental Psychology* 17, 3 (1963), 274–290.
Daniel E Berlyne. 1970. Novelty, complexity, and hedonic value. *Perception & Psychophysics* 8, 5 (1970), 279–286.
Daniel E Berlyne. 1971. *Aesthetics and psychobiology*. Vol. 336. JSTOR.
Eloïse Berson, Claire-Hélène Demarty, and Ngoc Q. K. Duong. 2017. Multimodality and Deep Learning when predicting Media Interestingness. In *MediaEval 2017 Workshop*.
Bir Bhanu. 1986. Automatic target recognition: State of the art survey. *IEEE transactions on aerospace and electronic systems* AES-22, 4 (1986), 364–379.
George David Birkhoff. 1933. *Aesthetic measure*. Vol. 38. Harvard University Press.
Kelsey Blackburn and James Schirillo. 2012. Emotive hemispheric differences measured in real-life portraits using pupil diameter and subjective aesthetic preferences. *Experimental brain research* 219, 4 (2012), 447–455.
Ali Borji, Ming-Ming Cheng, Huaizu Jiang, and Jia Li. 2015. Salient object detection: A benchmark. *IEEE transactions on image processing* 24, 12 (2015), 5706–5722.
Ali Borji and Laurent Itti. 2013. State-of-the-art in visual attention modeling. *IEEE transactions on pattern analysis and machine intelligence* 35, 1 (2013), 185–207.
Timothy F Brady, Talia Konkle, George A Alvarez, and Aude Oliva. 2008. Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Sciences* 105, 38 (2008), 14325–14329.
Markus M Breunig, Hans-Peter Kriegel, Raymond T Ng, and Jörg Sander. 2000. LOF: identifying density-based local outliers. In *ACM SIGMOD record*, Vol. 29. 93–104.
William H Burnham. 1908. Attention and interest. *The American Journal of Psychology* 19, 1 (1908), 14–18.
Daniel Stewart Butterfield, Caterina Fake, Callum James Henderson-Begg, and Serguei Mourachov. 2014. Interestingness ranking of media objects. Google Patents. US Patent 8,732,175.
Zoya Bylinskii, Michelle A Borkin, Nam Wook Kim, Hanspeter Pfister, and Aude Oliva. 2015a. Eye fixation metrics for large scale evaluation and comparison of information visualizations. In *Workshop on Eye Tracking and Visualization*. Springer, 235–255.
Zoya Bylinskii, Phillip Isola, Constance Bainbridge, Antonio Torralba, and Aude Oliva. 2015b. Intrinsic and extrinsic effects on image memorability. *Vision research* 116 (2015), 165–178.
Zoya Bylinskii, Adrià Recasens, Ali Borji, Aude Oliva, Antonio Torralba, and Frédo Durand. 2016. Where should saliency models look next?. In *European Conference on Computer Vision*. Springer, 809–824.
Michel Cabanac. 2002. What is emotion? *Behavioural processes* 60, 2 (2002), 69–83.
Maurizio Cardaci, Vito Di Gesù, Maria Petrou, and Marco Elio Tabacchi. 2009. A fuzzy approach to the evaluation of image complexity. *Fuzzy Sets and Systems* 160, 10 (2009), 1474–1484.
Rodrigo Andrés Cárdenas and Lauren Julius Harris. 2006. Symmetrical decorations enhance the attractiveness of faces and abstract designs. *Evolution and Human Behavior* 27, 1 (2006), 1–18.
Pew Research Center. 2016. News use across social media platforms 2016. (2016).
Souad Chaabouni, Jenny Benois-Pineau, Akka Zemmari, and Chokri Ben Amar. 2017. Deep saliency: prediction of interestingness in video with CNN. In *Visual Content Indexing and Retrieval with Psycho-Visual Models*. Springer, 43–74.
Christel Chamaret, Claire-Helene Demarty, Vincent Demoulin, and Gwenaelle Marquant. 2016. Experiencing the interestingness concept within and between pictures. *Electronic Imaging* 2016, 16 (2016), 1–12.
Arjun Chandrasekaran, Ashwin K Vijayakumar, Stanislaw Antol, Mohit Bansal, Dhruv Batra, C Lawrence Zitnick, and Devi Parikh. 2016. We are humor beings: Understanding and predicting visual humor. In *IEEE Conference on Computer Vision and Pattern Recognition*. 4603–4612.
Ang Chen, Paul W Darst, and Robert P Pangrazi. 2001. An examination of situational interest and its sources. *British Journal of Educational Psychology* 71, 3 (2001), 383–400.
Xinpeng Chen, Jingyuan Chen, Lin Ma, Jian Yao, Wei Liu, Jiebo Luo, and Tong Zhang. 2018. Fine-grained Video Attractiveness Prediction Using Multimodal Deep Learning on a Large Real-world Dataset. In *WWW 18 Companion: The 2018 Web Conference Companion*. 671–678.
Seo Young Choi, M Luo, Michael Pointer, and Peter Rhodes. 2009. Investigation of large display color image appearance–III: Modeling image naturalness. *Journal of Imaging Science and Technology* 53, 3 (2009), 31104–1.
Sharon Lynn Chu, Elena Fedorovskaya, Francis Quek, and Jeffrey Snyder. 2013. The effect of familiarity on perceived interestingness of images. In *IS&T/SPIE Electronic Imaging*, Vol. 8651. International Society for Optics and Photonics, 86511C–86511C.
Romain Cohendet, Karthik Yadati, Ngoc QK Duong, and Claire-Hélène Demarty. 2018. Annotating, Understanding, and Predicting Long-term Video Memorability. In *Proceedings of the 2018 ACM on International Conference on Multimedia Retrieval*. ACM, 178–186.
Mihai Gabriel Constantin, Bogdan Boteanu, and Bogdan Ionescu. 2016. LAPI at MediaEval 2016: Predicting Media Interestingness Task. In *MediaEval 2016 Workshop*.

Mihai Gabriel Constantin, Bogdan Boteanu, and Bogdan Ionescu. 2017. LAPI at MediaEval 2017: Predicting Media Interestingness. In *MediaEval 2017 Workshop*.

Gerald C Cupchik and Robert J Gebotys. 1990. Interest and pleasure as dimensions of aesthetic response. *Empirical Studies of the Arts* 8, 1 (1990), 1–14.

Ritendra Datta, Dhiraj Joshi, Jia Li, and James Z Wang. 2006. Studying aesthetics in photographic images using a computational approach. In *European Conference on Computer Vision*. Springer, 288–301.

Ritendra Datta, Jia Li, and James Z Wang. 2008. Algorithmic inferencing of aesthetics and emotion in natural images: An exposition. In *IEEE International Conference on Image Processing*. IEEE, 105–108.

Dmitry Davidov, Oren Tsur, and Ari Rappoport. 2010. Semi-supervised recognition of sarcastic sentences in twitter and amazon. In *Conference on Computational Natural Language Learning*. Association for Computational Linguistics, 107–116.

Claire-Helène Demarty, Mats Sjöberg, Mihai Gabriel Constantin, Ngoc QK Duong, Bogdan Ionescu, Thanh-Toan Do, and Hanli Wang. 2017a. Predicting Interestingness of Visual Content. In *Visual Content Indexing and Retrieval with Psycho-Visual Models*. Springer, 233–265.

Claire-Hélène Demarty, Mats Sjöberg, Bogdan Ionescu, Thanh-Toan Do, Michael Gygli, and Ngoc QK Duong. 2017b. MediaEval 2017 Predicting Media Interestingness Task. In *MediaEval 2017 Workshop*.

Claire-Hélène Demarty, Mats Sjöberg, Bogdan Ionescu, Thanh-Toan Do, Hanli Wang, Ngoc QK Duong, and Frédéric Lefebvre. 2016. Mediaeval 2016 predicting media interestingness task. In *MediaEval 2016 Workshop*.

Arturo Deza and Devi Parikh. 2015. Understanding image virality. In *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 1818–1826.

Sagnik Dhar, Vicente Ordonez, and Tamara L Berg. 2011. High level describable attributes for predicting aesthetics and interestingness. In *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 1657–1664.

Rachit Dubey, Joshua Peterson, Aditya Khosla, Ming-Hsuan Yang, and Bernard Ghanem. 2015. What makes an object memorable?. In *IEEE International Conference on Computer Vision*. IEEE, 1089–1097.

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

Lior Elazary and Laurent Itti. 2008. Interesting objects are visually salient. *Journal of vision* 8, 3 (2008), 3–3.

Phoebe C Ellsworth and Klaus R Scherer. 2003. Appraisal processes in emotion. *Handbook of affective sciences* 572 (2003), V595.

Phoebe C Ellsworth and Craig A Smith. 1988. Shades of joy: Patterns of appraisal differentiating pleasant emotions. *Cognition & Emotion* 2, 4 (1988), 301–331.

Shaojing Fan, Tian-Tsong Ng, Bryan L Koenig, Ming Jiang, and Qi Zhao. 2016. A Paradigm for Building Generalized Models of Human Image Perception through Data Fusion. In *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 5762–5771.

Kirill Fayn, Carolyn MacCann, Niko Tiliopoulos, and Paul J Silvia. 2015. Aesthetic emotions and aesthetic people: Openness predicts sensitivity to novelty in the experiences of interest and pleasure. *Frontiers in psychology* 6 (2015), 1877.

Mengjuan Fei, Wei Jiang, and Weijie Mao. 2018. Creating memorable video summaries that satisfy the users intention for taking the videos. *Neurocomputing* 275 (2018), 1911–1920.

Otto Fenichel. 1951. On the psychology of boredom. *Organization and pathology of thought* (1951), 349–361.

Catrin Finkenauer, Rutger CME Engels, and Wim Meeus. 2002. Keeping secrets from parents: Advantages and disadvantages of secrecy in adolescence. *Journal of Youth and Adolescence* 31, 2 (2002), 123–136.

John R Firth. 1957. A synopsis of linguistic theory, 1930-1955. (1957).

Johnny RJ Fontaine, Klaus R Scherer, Etienne B Roesch, and Phoebe C Ellsworth. 2007. The world of emotions is not two-dimensional. *Psychological science* 18, 12 (2007), 1050–1057.

Giulia Fontanini, Marco Bertini, and Alberto Del Bimbo. 2016. Web Video Popularity Prediction using Sentiment and Content Visual Features. In *ACM International Conference on Multimedia Retrieval*. ACM, 289–292.

Harry Fowler. 1965. *Curiosity and exploratory behavior*. Macmillan.

Barbara L. Fredrickson. 2001. *The role of positive emotions in positive psychology: The broaden-and-build theory of positive emotions.* Vol. 56. American Psychological Association. 218–226 pages.

Cassandra M Germain and Thomas M Hess. 2007. Motivational influences on controlled processing: Moderating distractibility in older adults. *Aging, Neuropsychology, and Cognition* 14, 5 (2007), 462–486.

Ross Girshick. 2015. Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision*. 1440–1448.

Helmut Grabner, Fabian Nater, Michel Druey, and Luc Van Gool. 2013. Visual interestingness in image sequences. In *ACM international conference on Multimedia*. ACM, 1017–1026.

Tom Grill and Mark Scanlon. 1990. *Photographic composition*. Amphoto Books.

Charles R. Gruner. 1970. The effect of humor in dull and interesting informative speeches. *Central States Speech Journal* 21, 3 (1970).

Marco Guerini, Jacopo Staiano, and Davide Albanese. 2013. Exploring image virality in google plus. In

International Conference on Social Computing (SocialCom). IEEE, 671–678.
Michael Gygli, Helmut Grabner, Hayko Riemenschneider, Fabian Nater, and Luc Van Gool. 2013. The Interestingness of Images. In *IEEE International Conference on Computer Vision*. IEEE, 1633–1640.
Michael Gygli, Helmut Grabner, Hayko Riemenschneider, and Luc Van Gool. 2014. Creating summaries from user videos. In *European conference on computer vision*. Springer, 505–520.
Michael Gygli, Helmut Grabner, and Luc Van Gool. 2015. Video summarization by learning submodular mixtures of objectives. In *IEEE Conference on Computer Vision and Pattern Recognition*. 3090–3098.
Michael Gygli and Mohammad Soleymani. 2016. Analyzing and predicting GIF interestingness. In *ACM on Multimedia Conference*. ACM, 122–126.
Raisa Halonen, Stina Westman, and Pirkko Oittinen. 2011. Naturalness and interestingness of test images for visual quality evaluation. In *IS&T/SPIE Electronic Imaging*. International Society for Optics and Photonics, 78670Z–78670Z.
R Hamed, Adham Atyabi, Antti Rantanen, Seppo J Laukka, Samia Nefti-Meziani, Janne Heikkilä, and others. 2015. Predicting the valence of a scene from observers eye movements. *PloS one* 10, 9 (2015), e0138198.
Robert M Haralick. 1979. Statistical and structural approaches to texture. *Proc. IEEE* 67, 5 (1979), 786–804.
Ralf Herbrich, Tom Minka, and Thore Graepel. 2006. TrueSkill: a Bayesian skill rating system. In *International Conference on Neural Information Processing Systems*. 569–576.
Alejandro Hernández-García, Fernando Fernández-Martínez, and Fernando Díaz-de María. 2016. Comparing visual descriptors and automatic rating strategies for video aesthetics prediction. *Signal Processing: Image Communication* 47 (2016), 280–288.
Eckhard H Hess and James M Polt. 1960. Pupil size as related to interest value of visual stimuli. *Science* 132, 3423 (1960), 349–350.
Suzanne Hidi and Valerie Anderson. 1992. Situational interest and its impact on reading and expository writing. *The role of interest in learning and development* 11 (1992), 213–214.
Suzanne E Hidi. 1995. A reexamination of the role of attention in learning from text. *Educational Psychology Review* 7, 4 (1995), 323–350.
James E Hoffman and Baskaran Subramaniam. 1995. The role of visual attention in saccadic eye movements. *Perception & psychophysics* 57, 6 (1995), 787–795.
Liang-Chi Hsieh, Winston H Hsu, and Hao-Chuan Wang. 2014. Investigating and predicting social and visual image interestingness on social media by crowdsourcing. In *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 4309–4313.
Kai-Qi Huang, Qiao Wang, and Zhen-Yang Wu. 2006. Natural color image enhancement and evaluation algorithm based on human visual system. *Computer Vision and Image Understanding* 103, 1 (2006), 52–63.
David A Huffman. 1952. A method for the construction of minimum-redundancy codes. *Proceedings of the IRE* 40, 9 (1952), 1098–1101.
R Reed Hunt and James B Worthen. 2006. *Distinctiveness and memory*. Oxford University Press.
Phillip Isola, Devi Parikh, Antonio Torralba, and Aude Oliva. 2011a. Understanding the intrinsic memorability of images. In *Advances in Neural Information Processing Systems*. 2429–2437.
Phillip Isola, Jianxiong Xiao, Torralba Antonio, and Aude Oliva. 2011b. What makes an image memorable?. In *Conference on Computer Vision and Pattern Recognition*. IEEE, 145–152.
Phillip Isola, Jianxiong Xiao, Devi Parikh, Antonio Torralba, and Aude Oliva. 2014. What Makes a Photograph Memorable? *IEEE Transactions on Pattern Analysis and Machine Intelligence* 36, 7 (2014), 1469–1482.
C. E. Izard and B. P. Ackerman. 2010. Emotion-cognition relationships and human development. *Lewis, Michael and Haviland-Jones, Jeannette M, Handbook of emotions* (2010), 253–264.
Philip W Jackson and Samuel Messick. 1965. The person, the product, and the response: conceptual problems in the assessment of creativity1. *Journal of personality* 33, 3 (1965), 309–329.
TJWM Janssen and FJJ Blommaert. 2000. Predicting the usefulness and naturalness of color reproductions. *Journal of imaging science and Technology* 44, 2 (2000), 93–104.
Yu-Gang Jiang, Yanran Wang, Rui Feng, Xiangyang Xue, Yingbin Zheng, and Hanfang Yang. 2013. Understanding and Predicting Interestingness of Videos. In *Proceedings of the Twenty-Seventh AAAI Conference on Artificial Intelligence*. ACM, AAAI Press, 1113–1119.
Yu-Gang Jiang, Baohan Xu, and Xiangyang Xue. 2014. Predicting Emotions in User-Generated Videos. In *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence*. ACM, 73–79.
Anna Jordanous. 2012. A standardised procedure for evaluating creative systems: Computational creativity evaluation based on what it is to be creative. *Cognitive Computation* 4, 3 (2012), 246–279.
Brendan Jou, Subhabrata Bhattacharya, and Shih-Fu Chang. 2014. Predicting viewer perceived emotions in animated GIFs. In *ACM international conference on Multimedia*. ACM, 213–216.
Brendan Jou, Tao Chen, Nikolaos Pappas, Miriam Redi, Mercan Topkara, and Shih-Fu Chang. 2015. Visual affect around the world: A large-scale multilingual visual sentiment ontology. In *ACM international conference on Multimedia*. ACM, 159–168.

Immanuel Kant and Werner S Pluhar. 1987. *Critique of judgment*. Hackett Publishing.
Yan Ke, Xiaoou Tang, and Feng Jing. 2006. The design of high-level features for photo quality assessment. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. 1. IEEE, 419–426.
Aditya Khosla, Akhil S Raju, Antonio Torralba, and Aude Oliva. 2015. Understanding and predicting image memorability at a large scale. In *IEEE International Conference on Computer Vision*. IEEE, 2390–2398.
Aditya Khosla, Atish Das Sarma, and Raffay Hamid. 2014. What makes an image popular?. In *International conference on World wide web*. ACM, 867–876.
Aditya Khosla, Jianxiong Xiao, Antonio Torralba, and Aude Oliva. 2012. Memorability of image regions. In *Advances in Neural Information Processing Systems*. Elsevier, 305–313.
Sander Koelstra, Christian Muhl, Mohammad Soleymani, Jong-Seok Lee, Ashkan Yazdani, Touradj Ebrahimi, Thierry Pun, Anton Nijholt, and Ioannis Patras. 2012. Deap: A database for emotion analysis; using physiological signals. *IEEE Transactions on Affective Computing* 3, 1 (2012), 18–31.
Kurt Koffka. 2013. *Principles of Gestalt psychology*. Routledge.
Bert Krages. 2012. *Photography: the art of composition*. Skyhorse Publishing, Inc.
Roger J Kreuz and Sam Glucksberg. 1989. How to be sarcastic: The echoic reminder theory of verbal irony. *Journal of Experimental Psychology: General* 118, 4 (1989), 374–386.
Vu Lam, Tien Do, Sang Phan, Duy-Dinh Le, Shinichi Satoh, and Duc Anh Duong. 2016. NII-UIT at MediaEval 2016 Predicting Media Interestingness Task. In *MediaEval 2016 Workshop*.
Peter J Lang, Margaret M Bradley, and Bruce N Cuthbert. 1999. International affective picture system (IAPS): Instruction manual and affective ratings. *The center for research in psychophysiology, University of Florida* (1999).
Svetlana Lazebnik, Cordelia Schmid, and Jean Ponce. 2006. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In *Conference on Computer Vision and Pattern Recognition*. IEEE, 2169–2178.
Olivier Le Meur, Thierry Baccino, and Aline Roumy. 2011. Prediction of the inter-observer visual congruency (IOVC) and application to image ranking. In *Proceedings of the 19th ACM international conference on Multimedia*. ACM, 373–382.
Jieun Lee and Ilyoo B Hong. 2016. Predicting positive user responses to social media advertising: The roles of emotional appeal, informativeness, and creativity. *International Journal of Information Management* 36, 3 (2016), 360–373.
Li-Jia Li, Hao Su, Li Fei-Fei, and Eric P Xing. 2010. Object bank: A high-level image representation for scene classification & semantic feature sparsification. In *Advances in neural information processing systems*. Curran Associates, Inc., 1378–1386.
Ming Li and Paul Vitányi. 2009. *An introduction to Kolmogorov complexity and its applications*. Springer Science & Business Media.
Cynthia CS Liem. 2016. TUD-MMC at MediaEval 2016: Predicting Media Interestingness Task. In *MediaEval 2016 Workshop*.
Annukka K Lindell and Julia Mueller. 2011. Can science account for taste? Psychological insights into art appreciation. *Journal of Cognitive Psychology* 23, 4 (2011), 453–475.
Feng Liu, Yuzhen Niu, and Michael Gleicher. 2009. Using Web Photos for Measuring Video Frame Interestingness.. In *Twenty-First International Joint Conference on Artificial Intelligence*. 2058–2063.
Tie Liu, Zejian Yuan, Jian Sun, Jingdong Wang, Nanning Zheng, Xiaoou Tang, and Heung-Yeung Shum. 2011. Learning to detect a salient object. *IEEE Transactions on Pattern analysis and machine intelligence* 33, 2 (2011), 353–367.
Yang Liu, Zhonglei Gu, Tobey H Ko, and Kien A Hua. 2018. Learning Perceptual Embeddings with Two Related Tasks for Joint Predictions of Media Interestingness and Emotions. In *Proceedings of the 2018 ACM on International Conference on Multimedia Retrieval*. ACM, 420–427.
Paul Locher, Els Cornelis, Johan Wagemans, and Pieter Jan Stappers. 2001. Artists' use of compositional balance for creating visual displays. *Empirical studies of the arts* 19, 2 (2001), 213–227.
Edward Harrington Lockwood and Robert Hugh Macmillan. 1978. *Geometric symmetry*. CUP Archive.
Konrad Lorenz. 1971. *Studies in animal and human behavior*. Methuen & Company Limited.
Konrad Lorenz and Marjorie Kerr Wilson. 2002. *Man meets dog*. Psychology Press.
Yiwen Luo and Xiaoou Tang. 2008. Photo and video quality evaluation: Focusing on the subject. In *European Conference on Computer Vision*. Springer, 386–399.
Kevin Lynch. 1960. *The image of the city*. Vol. 11. MIT press.
Jana Machajdik and Allan Hanbury. 2010. Affective image classification using features inspired by psychology and art theory. In *ACM international conference on Multimedia*. ACM, 83–92.
Mary Lou Maher. 2010. Evaluating creativity in humans, computers, and collectively intelligent systems. In *Conference on Creativity and Innovation in Design*. Desire Network, 22–28.
Frank H Mahnke. 1996. *Color, environment, and human response: an interdisciplinary understanding of color and its use as a beneficial element in the design of the architectural environment*. John Wiley & Sons.
Long Mai, Hailin Jin, and Feng Liu. 2016. Composition-preserving deep photo aesthetics assessment. In

IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 497–506.
Matei Mancas and Olivier Le Meur. 2013. Memorability of natural scenes: The role of attention. In *Image Processing (ICIP), 2013 20th IEEE International Conference on*. IEEE, 196–200.
Markos Markou and Sameer Singh. 2003. Novelty detection: a reviewpart 1: statistical approaches. *Signal processing* 83, 12 (2003), 2481–2497.
Gwenaelle Marquant, Claire-Hélène Demarty, Christel Chamaret, Joël Sirot, and Louis Chevallier. 2018. Interestingness Prediction & its Application to Immersive Content. In *2018 International Conference on Content-Based Multimedia Indexing (CBMI)*. IEEE, 1–6.
Rod A Martin, Patricia Puhlik-Doris, Gwen Larsen, Jeanette Gray, and Kelly Weir. 2003. Individual differences in uses of humor and their relation to psychological well-being: Development of the Humor Styles Questionnaire. *Journal of research in personality* 37, 1 (2003), 48–75.
W. Cheyne McCallum. 1999. Encyclopaedia Britannica. https://www.britannica.com/topic/attention
Mark A McDaniel, Paula J Waddill, Kraig Finstad, and Tammy Bourg. 2000. The effects of text-based interest on attention and recall. *Journal of Educational Psychology* 92, 3 (2000), 492.
A Peter McGraw, Caleb Warren, Lawrence E Williams, and Bridget Leonard. 2012. Too close for comfort, or too far to care? Finding humor in distant tragedies and close mishaps. *Psychological Science* 23, 10 (2012), 1215–1223.
Joseph A Mikels, Barbara L Fredrickson, Gregory R Larkin, Casey M Lindberg, Sam J Maglio, and Patricia A Reuter-Lorenz. 2005. Emotional category data on images from the International Affective Picture System. *Behavior research methods* 37, 4 (2005), 626–630.
William L Mikulas and Stephen J Vodanovich. 1993. The essence of boredom. *The Psychological Record* 43, 1 (1993), 3.
Shasha Mo, Jianwei Niu, Yiming Su, and Sajal K Das. 2018. A novel feature set for video emotion recognition. *Neurocomputing* 291 (2018), 11–20.
Douglas Colin Muecke and DC Muecke. 1969. *The compass of irony*. Oxford Univ Press.
Matthijs P Mulder and Antinus Nijholt. 2002. Humour research: State of art. (2002).
Naila Murray, Luca Marchesotti, and Florent Perronnin. 2012. AVA: A large-scale database for aesthetic visual analysis. In *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2408–2415.
Karen Nelson-Field, Erica Riebe, and Kellie Newstead. 2013. The emotions that drive viral video. *Australasian Marketing Journal (AMJ)* 21, 4 (2013), 205–211.
Pardis Noorzad and Bob L Sturm. 2012. Regression with sparse approximations of data. In *European Signal Processing Conference (EUSIPCO)*. IEEE, 674–678.
Timo Ojala, Matti Pietikainen, and Topi Maenpaa. 2002. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on pattern analysis and machine intelligence* 24, 7 (2002), 971–987.
Aude Oliva and Antonio Torralba. 2001. Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision* 42, 3 (2001), 145–175.
Selim Onat, Alper Açık, Frank Schumann, and Peter König. 2014. The contributions of image content and behavioral relevancy to overt attention. *PloS One* 9, 4 (2014), e93254.
Balaji Padmanabhan and Alexander Tuzhilin. 1999. Unexpectedness as a measure of interestingness in knowledge discovery. *Decision Support Systems* 27, 3 (1999), 303–318.
Jayneel Parekh, Harshvardhan Tibrewal, and Sanjeel Parekh. 2017. The IITB Predicting Media Interestingness System for MediaEval 2017. In *MediaEval 2017 Workshop*.
Jayneel Parekh, Harshvardhan Tibrewal, and Sanjeel Parekh. 2018. Deep Pairwise Classification and Ranking for Predicting Media Interestingness. In *Proceedings of the 2018 ACM on International Conference on Multimedia Retrieval*. ACM, 428–433.
Jennifer T Parkhurst and Andrea Hopmeyer. 1998. Sociometric popularity and peer-perceived popularity two distinct dimensions of peer status. *The Journal of Early Adolescence* 18, 2 (1998), 125–144.
Kuan-Chuan Peng, Tsuhan Chen, Amir Sadovnik, and Andrew C Gallagher. 2015. A mixed bag of emotions: Model, predict, and transfer emotion distributions. In *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 860–868.
Jukka Perkiö and Aapo Hyvärinen. 2009. Modelling image complexity by independent component analysis, with application to content-based image retrieval. *International Conference on Artificial Neural Networks* (2009), 704–714.
Reza Aditya Permadi, Septian Gilang Permana Putra, Helmiriawan, and Cynthia C. S. Liem. 2017. DUT-MMSR at MediaEval 2017: Predicting Media Interestingness Task. In *MediaEval 2017 Workshop*.
Richard Alan Peters and Robin N Strickland. 1990. Image complexity metrics for automatic target recognizers. In *Automatic Target Recognizer System and Technology Conference*. 1–17.
WA Phillips. 1974. On the distinction between sensory storage and short-term visual memory. *Perception & Psychophysics* 16, 2 (1974), 283–290.
Robert Plutchik. 1980. A general psychoevolutionary theory of emotion. *Theories of emotion* 1 (1980), 3–31.
John P Powell and Lee W Andresen. 1985. Humour and teaching in higher education. *Studies in Higher Education* 10, 1 (1985), 79–90.

Daniele Quercia, Rossano Schifanella, and Luca Maria Aiello. 2014. The shortest path to happiness: Recommending beautiful, quiet, and happy routes in the city. In *ACM conference on hypertext and social media*. ACM, 116–125.

Rolf Reber, Norbert Schwarz, and Piotr Winkielman. 2004. Processing fluency and aesthetic pleasure: Is beauty in the perceiver's processing experience? *Personality and social psychology review* 8, 4 (2004), 364–382.

Miriam Redi and Bernard Merialdo. 2012. Where is the interestingness? Retrieving appealing video scenes by learning Flickr-based graded judgments. In *International Conference on Multimedia Retrieval*. ACM, 1363–1364.

Miriam Redi, Neil O'Hare, Rossano Schifanella, Michele Trevisiol, and Alejandro Jaimes. 2014. 6 seconds of sound and vision: Creativity in micro-videos. In *Conference on Computer Vision and Pattern Recognition*. IEEE, 4272–4279.

Miriam Redi, Nikhil Rasiwasia, Gaurav Aggarwal, and Alejandro Jaimes. 2015. The Beauty of Capturing Faces: Rating the Quality of Digital Portraits. In *IEEE International Conference on Automatic Face and Gesture Recognition*, Vol. 1. IEEE, 1–8.

Ronald A Rensink, J Kevin O'Regan, and James J Clark. 1997. To see or not to see: The need for attention to perceive changes in scenes. *Psychological science* 8, 5 (1997), 368–373.

Mel Rhodes. 1961. An analysis of creativity. *The Phi Delta Kappan* 42, 7 (1961), 305–310.

Jaume Rigau, Miquel Feixas, and Mateu Sbert. 2007. Conceptualizing Birkhoff's Aesthetic Measure Using Shannon Entropy and Kolmogorov Complexity. In *Computational Aesthetics*. Eurographics Association, 105–112.

Ellen Riloff, Ashequl Qadir, Prafulla Surve, Lalindra De Silva, Nathan Gilbert, and Ruihong Huang. 2013. Sarcasm as Contrast between a Positive Sentiment and Negative Situation. In *Conference on Empirical Methods in Natural Language Processing*, Vol. 13. ACL, 704–714.

Graeme Ritchie. 1999. Developing the Incongruity-Resolution Theory. *Institute for Communicating and Collaborative Systems* (1999).

Daniel M. Romero, Wojciech Galuba, Sitaram Asur, and Bernardo A. Huberman. 2011. Influence and Passivity in Social Media. In *Machine Learning and Knowledge Discovery in Databases*, Dimitrios Gunopulos, Thomas Hofmann, Donato Malerba, and Michalis Vazirgiannis (Eds.). Lecture Notes in Computer Science, Vol. 6913. Springer Berlin Heidelberg, 18–33.

Ira J Roseman. 1996. Appraisal determinants of emotions: Constructing a more accurate and comprehensive theory. *Cognition & Emotion* 10, 3 (1996), 241–278.

Ruth Rosenholtz, Yuanzhen Li, and Lisa Nakano. 2007. Measuring visual clutter. *Journal of vision* 7, 2 (2007), 17–17.

Yossi Rubner, Carlo Tomasi, and Leonidas J Guibas. 1998. A metric for distributions with applications to image databases. In *Computer Vision, 1998. Sixth International Conference on*. IEEE, 59–66.

Willibald Ruch. 2008. Psychology of humor. *A primer of humor* (2008), 17–100.

PA Russell and DA George. 1990. Relationships between aesthetic response scales applied to paintings. *Empirical Studies of the Arts* 8, 1 (1990), 15–30.

Jonathan Sammartino and Stephen E Palmer. 2012. Aesthetic issues in spatial composition: Effects of vertical position and perspective on framing single objects. *Journal of Experimental Psychology: Human Perception and Performance* 38, 4 (2012), 865.

Darshan Santani, Salvador Ruiz-Correa, and Daniel Gatica-Perez. 2017. Insiders and Outsiders: Comparing Urban Impressions between Population Groups. In *ACM on International Conference on Multimedia Retrieval*. ACM, 65–71.

Andreza Sartori, Dubravko Culibrk, Yan Yan, and Nicu Sebe. 2015a. Who's Afraid of Itten: Using the Art Theory of Color Combination to Analyze Emotions in Abstract Paintings. In *ACM international conference on Multimedia*. ACM, 311–320.

Andreza Sartori, Victoria Yanulevskaya, Almila Akdag Salah, Jasper Uijlings, Elia Bruni, and Nicu Sebe. 2015b. Affective analysis of professional and amateur abstract paintings using statistical analysis and art theory. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 5, 2 (2015), 8:1–8:27.

Rossano Schifanella, Miriam Redi, and Luca Maria Aiello. 2015. An Image Is Worth More than a Thousand Favorites: Surfacing the Hidden Beauty of Flickr Pictures. In *AAAI Conference on Web and Social Media*.

Karen B Schloss, Eli D Strauss, and Stephen E Palmer. 2013. Object color preferences. *Color Research & Application* 38, 6 (2013), 393–411.

Jürgen Schmidhuber. 2009. Driven by compression progress: A simple principle explains essential aspects of subjective beauty, novelty, surprise, interestingness, attention, curiosity, creativity, art, science, music, jokes. In *Anticipatory Behavior in Adaptive Learning Systems*. Springer, 48–76.

M Shapely. 1987. Prosodic variation and audience response. *IPrA: Papers in Pragmatics* 1, 2 (1987), 66–79.

Sumit Shekhar, Dhruv Singal, Harvineet Singh, Manav Kedia, and Akhil Shetty. 2017. Show and Recall: Learning What Makes Videos Memorable. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2730–2739.

Yuesong Shen, Claire-Hélène Demarty, and Ngoc QK Duong. 2016. Technicolor@ MediaEval 2016 Predicting

Media Interestingness Task. In *MediaEval 2016 Workshop*.
Roger N. Shepard. 1967. Recognition memory for words, sentences, and pictures. *Journal of verbal Learning and verbal Behavior* 6, 1 (1967), 156–163.
Yue Shi and Martha Larson. 2010. First Approaches to Automatic Boredom Detection: DMIR tackles the MediaEval 2010 Affect Task. In *MediaEval 2010 Workshop*.
Larry L Shirey and Ralph E Reynolds. 1988. Effect of interest on attention and learning. *Journal of Educational Psychology* 80, 2 (1988), 159.
Aliaksandr Siarohin, Gloria Zen, Cveta Majtanovic, Xavier Alameda-Pineda, Elisa Ricci, and Nicu Sebe. 2017. How to Make an Image More Memorable? A Deep Style Transfer Approach. In *International Conference on Multimedia Retrieval*. ACM, 322–329.
Paul J Silvia. 2005. What is interesting? Exploring the appraisal structure of interest. *Emotion* 5, 1 (2005), 89–102.
Paul J Silvia. 2006. *Exploring the psychology of interest*. Oxford University Press.
Paul J Silvia. 2008. Interest-The curious emotion. In *Current Directions in Psychological Science*, Vol. 17. 57–60.
Paul J Silvia. 2009. Looking past pleasure: Anger, confusion, disgust, pride, surprise, and other unusual aesthetic emotions. *Psychology of Aesthetics, Creativity, and the Arts* 3, 1 (2009), 48.
Paul J Silvia and John B Warburton. 2006. Positive and negative affect: Bridging states and traits. *Comprehensive handbook of personality and psychopathology* 1 (2006), 268–284.
Mohammad Soleymani. 2010. Travelogue Boredom Detection with Content Features. In *MediaEval 2010 Workshop*.
Mohammad Soleymani. 2015. The quest for visual interest. In *ACM international conference on Multimedia*. ACM, 919–922.
Mohammad Soleymani, Guillaume Chanel, Joep JM Kierkels, and Thierry Pun. 2008. Affective characterization of movie scenes based on multimedia content analysis and user's physiological emotional responses. In *International Symposium on Multimedia (ISM)*. IEEE, 228–235.
Duncan Southgate, Nikki Westoby, and Graham Page. 2010. Creative determinants of viral video viewing. *International Journal of Advertising* 29, 3 (2010), 349–368.
Angus Stevenson. 2010. *Oxford dictionary of English*. Oxford University Press, USA.
Andreas Stolcke, Xavier Anguera, Kofi Boakye, Özgür Çetin, Adam Janin, Mathew Magimai-Doss, Chuck Wooters, and Jing Zheng. 2008. The SRI-ICSI Spring 2007 meeting and lecture recognition system. In *Multimodal Technologies for Perception of Humans*. Springer, 450–463.
Masaki Takahashi and Masanori Sano. 2013. NHK Where is beauty? Grand Challenge. In *ACM Multimedia challenge*. http://acmmm13.org/submissions/call-for-multimedia-grand-challenge-solutions/task-where-is-beauty/
Silvan S Tomkins. 1962. Affect, imagery, consciousness: Vol. I. The positive affects. (1962).
Lorenzo Torresani, Martin Szummer, and Andrew Fitzgibbon. 2010. Efficient object category recognition using classemes. In *European conference on computer vision*. Springer, 776–789.
Du Tran, Lubomir Bourdev, Rob Fergus, Lorenzo Torresani, and Manohar Paluri. 2015. Learning spatiotemporal features with 3d convolutional networks. In *IEEE International Conference on Computer Vision*. IEEE, 4489–4497.
Samuel A Turner and Paul J Silvia. 2006. Must interesting things be pleasant? A test of competing appraisal structures. *Emotion* 6, 4 (2006), 670.
Hang Maxime Ung. 2011. Social Influence, Popularity and Interestingness of Online Contents. In *Fifth International AAAI Conference on Weblogs and Social Media*.
Akira Utsumi. 1996. Implicit display theory of verbal irony: Towards a computational model of irony. *Hulstijn and Nijholt* (1996), 29–38.
Patricia Valdez and Albert Mehrabian. 1994. Effects of color on emotions. *Journal of experimental psychology: General* 123, 4 (1994), 394–409.
Joost Van de Weijer, Cordelia Schmid, and Jakob Verbeek. 2007. Learning color names from real-world images. In *Conference on Computer Vision and Pattern Recognition*. IEEE, 1–8.
Mike Van Duuren, Linda Kendell-Scott, and Natalie Stark. 2003. Early aesthetic choices: Infant preferences for attractive premature infant faces. In *International Journal of Behavioral Development*, Vol. 27. 212–219.
Jagannadan Varadarajan and Jean-Marc Odobez. 2009. Topic models for scene analysis and abnormality detection. In *International Conference on Computer Vision Workshops*. IEEE, 1338–1345.
P-A Verhaegen, Dennis Vandevenne, and JR Duflou. 2012. Originality and Novelty: a different universe. In *International Design Conference*. 1961–1966.
Stine Vogt and Svein Magnussen. 2007. Long-term memory for 400 pictures on a common theme. *Experimental psychology* 54, 4 (2007), 298–303.
Shuai Wang, Shizhe Chen, Jinming Zhao, Wenxuan Wang, and Qin Jin. 2017. RUC at MediaEval 2017: Predicting Media Interestingness Task. In *MediaEval 2017 Workshop*.
Yanran Wang, Qi Dai, Rui Feng, and Yu-Gang Jiang. 2013. Beauty is here: evaluating aesthetics in videos

using multimodal features and free training data. In *International conference on Multimedia*. ACM, 369–372.

Amy Beth Warriner, Victor Kuperman, and Marc Brysbaert. 2013. Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior research methods* 45, 4 (2013), 1191–1207.

Robert W Weisberg. 1999. I2 Creativity and Knowledge: A Challenge to Theories. *Handbook of creativity* 226–250 (1999).

Lois B Wexner. 1954. The degree to which colors (hues) are associated with mood-tones. *Journal of applied Psychology* 38, 6 (1954), 432–435.

Michael J. Wilber, Chen Fang, Hailin Jin, Aaron Hertzmann, John Collomosse, and Serge Belongie. 2017. BAM! The Behance Artistic Media Dataset for Recognition Beyond Photography. In *IEEE International Conference on Computer Vision*. IEEE, 1211–1220.

Jianxiong Xiao, James Hays, Krista A Ehinger, Aude Oliva, and Antonio Torralba. 2010. Sun database: Large-scale scene recognition from abbey to zoo. In *IEEE conference on Computer Vision and Pattern Recognition*. IEEE, 3485–3492.

Baohan Xu, Yanwei Fu, and Yu-Gang Jiang. 2016. BigVid at MediaEval 2016: Predicting Interestingness in Images and Videos. In *MediaEval 2016 Workshop*.

Baohan Xu, Yanwei Fu, Yu-Gang Jiang, Boyang Li, and Leonid Sigal. 2018. Heterogeneous knowledge transfer in video emotion recognition, attribution and summarization. *IEEE Transactions on Affective Computing* 9, 2 (2018), 255–270.

Che-Hua Yeh, Yuan-Chen Ho, Brian A Barsky, and Ming Ouhyoung. 2010. Personalized photograph ranking and selection system. In *International conference on Multimedia*. ACM, 211–220.

Suet-Peng Yong, Jeremiah D Deng, and Martin K Purvis. 2013. Wildlife video key-frame extraction based on novelty detection in semantic context. *Multimedia Tools and Applications* 62, 2 (2013), 359–376.

Sejong Yoon. 2017. TCNJ-CS@MediaEval 2017 Predicting Media Interestingness Task. In *MediaEval 2017 Workshop*.

Sejong Yoon and Vladimir Pavlovic. 2014. Sentiment Flow for Video Interestingness Prediction. In *International Workshop on Human Centered Event Understanding from Multimedia*. ACM, 29–34.

Hui Yu, Mingjing Li, Hong-Jiang Zhang, and Jufu Feng. 2002. Color texture moments for content-based image retrieval. In *International Conference om Image Processing*, Vol. 3. IEEE, 929–932.

Honghai Yu and Stefan Winkler. 2013. Image complexity and spatial information. In *International Workshop on Quality of Multimedia Experience (QoMEX)*. IEEE, 12–17.

Nick Zangwill. 2003. Aesthetic Judgment. In *The Stanford encyclopedia of philosophy* (fall 2007 ed. ed.), E. N. Zalta (Ed.). http://plato.stanford.edu/entries/aesthetic-judgment/

Gloria Zen, Paloma de Juan, Yale Song, and Alejandro Jaimes. 2016. Mouse activity as an indicator of interestingness in video. In *International Conference on Multimedia Retrieval*. ACM, 47–54.

Bin Zhao, Li Fei-Fei, and Eric P Xing. 2011. Online detection of unusual events in videos via dynamic sparse coding. In *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 3313–3320.

Qi Zhao and Christof Koch. 2013. Learning saliency-based visual attention: A review. *Signal Processing* 93, 6 (2013), 1401–1407.

Sicheng Zhao, Yue Gao, Xiaolei Jiang, Hongxun Yao, Tat-Seng Chua, and Xiaoshuai Sun. 2014. Exploring principles-of-art features for image emotion recognition. In *ACM international conference on Multimedia*. ACM, 47–56.

Sicheng Zhao, Hongxun Yao, Yue Gao, Rongrong Ji, Wenlong Xie, Xiaolei Jiang, and Tat-Seng Chua. 2016. Predicting personalized emotion perceptions of social images. In *ACM Conference on Multimedia*. 1385–1394.

Bolei Zhou, Agata Lapedriza, Jianxiong Xiao, Antonio Torralba, and Aude Oliva. 2014. Learning deep features for scene recognition using places database. In *Advances in neural information processing systems*. 487–495.