Review of Methods to Predict Social Image Interestingness and Memorability

Conference Paper · September 2015

DOI: 10.1007/978-3-319-23192-1_6

Xesca Amengual

Universitat de Girona

Josep Lluis de la Rosa

Universitat de Girona

Review of Methods to Predict Social Image Interestingness and Memorability

Xesca Amengual, Anna Bosch, and Josep Lluís de la Rosa

DEEEA, Centre Easy, Agents Research LAB
Universitat de Giroan, Girona, Spain
{xesca.amengual,peplluis}@silver.udg.edu,annabosch@easyinnova.com

Abstract. An entire industry has developed around keyword optimization for ad buyers. However, social media landscape has shift to a photo driven behavior and there is a need to overcome the challenge to analyze all this large amount of visual data that users post in internet. We will address this analysis by providing a review on how to measure image and video interestingness and memorability from content that is tacked spontaneously in social networks. We will investigate current state-of-the-art of methods analyzing social media images and provide further research directions that could be beneficial for both, users and companies.

Keywords: interestingness, memorability, image, video, review

1 Introduction

The total number of internet users around the world was stated up to three billion in 2014[1] whereas the number of social network users worldwide from 2010 to 2014 has grown from 0.97 billion users to 1.79. In 2016, it is estimated that there will be around 2.13 billion social network users around the globe, up from 1.4 billion in 2012[2]. The amount of users generated data is huge and there is a need to provide tools to automatically process it.

To date, internet data and specifically social networks generated data have been monetized primarily by text-based applications. In fact, an entire industry has developed around keyword optimization for ad buyers, text analysis for opinion [1] and sentiment discovery [2], brand positioning [3], user behavior [4] and so on. Words drive economy of the web. However, it appears that Social Networks are now more show than tell. The shift to a personal newspaper-style format with larger and more prominent photo displays is a response to photo driven behavior that has rapidly changed the social media landscape. Machines that monetize the internet need to keep up with the times to analyze the large amount of uploaded photographs and video to internet.

The motivation of this work is twofold. First, it is well known that some images/videos get much more views than others so there is a need to understand

[1] http://www.internetlivestats.com/internet-users/

[2] http://www.statista.com/statistics/278414/

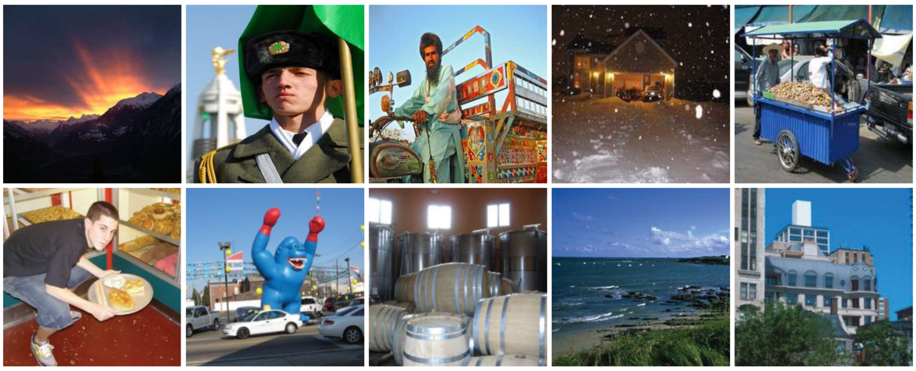

Fig. 1: Images ranked by their interestingness (top row) and memorability (bo0ttom row). Images are sortd form high (left) to low (right) score. [5, 6]

what makes an image or a video more interesting/popular/memorable[3] than others from a computer vision perspective (Fig. 1). Some works have recently started to address this theme with different techniques and methodologies, so a review to state best methodologies and further directions is needed. Second, text is no longer enough to provide monetization tools over the current internet behavior, there is a need to include knowledge from images in this process. We will provide here some clues on how this information could be considered. The rest of the paper is organized as follows. Since interestingness and memorability have different objectives, we provide, in Section 2 and 3, an overview of the most relevant research works respectively. The used datasets are detailed in Section 4. Finally, in Section 5 we give the conclusions and future work.

2 Interestingness

Interestingness is said to be the power of attracting or holding ones attention. This property has been object of study with several goals, such as knowledge discovery [7], association patterns [8] or Wikipedia data [9]. In this survey we focus on image and video interestingness. For years, psychological researchers have proposed several variables that affect interestingness measures. Berlyne [10] considered novelty, uncertainty, conflict and complexity, whereas Chen et al. [11] identified novelty, challenge, instant enjoyment and demand for attention as the most relevant cues. According to [12] high pleasantness was the major aspect of interestingness and [13] supported the presence of polygons and painting.

Although interestingness is clearly a subjective property and depends on personal preferences and experiences, there exists a significant agreement among users about which images are considered more interesting than others, and this has encouraged researches from psychology and now from computer vision to learn more about this topic. In addition, video interestingness research has also been addressed by the computer vision field for different applications such as video retrieval or video summarization by selecting the most interesting scenes.

[3] In this paper we consider interesting and popular as synonyms

2.1 State-of-the-art

To the best of our knowledge, the study of image interestingness is a novel research line and there are only a few papers addressing this topic from a computer vision perspective. Some studies compare their predictions with crowdsourcing results [14–16] whereas others prefer to compare them with the actual interestingness or popularity raised by images in social networks [5, 17–19]. In addition to image features, the usefulness of social cues is studied for the latter choice.

Grabner et al. [14] investigate how different features perform to determine which events are considered of interest in image sequences recorded by a static video camera. They highlight emotion (depending on brightness an saturation), complexity (bytes of encoded image) and novelty (outlier detection) of images, as well as, an interestingness score learned directly from gist features using a ν-SVR. Performance of individual features states that novelty is the best cue for this task, however best results are obtained when combining all of them by training a simple linear model with a ν-SVR. Even though the proposed cues are of interest, the context of image sequences plays an important role on the performed experiments and one may think that this method cannot be extended to other less contextualized image datasets.

In [15] authors study the correlation of interestingness with aesthetics and memorability. They find that interestingness is really correlated with aesthetics but, contrary to the popular belief, the correlation with memorability is very low. They train a ν-SVR to build a predictor using several cues, some of them taken from previous work [14] and other novel features for the purpose of interestingness prediction (see Table 1 for the specific features used). These attributes are selected to emphasize aspects such as unusualness, aesthetics and other general preferences. The method is applied over three datasets (the Webcam dataset (WD) used in [14] of strong context, the Scene Categories dataset (SCD) [20] of weak context and the Memorability dataset (MD) [6] of arbitrary photos) and predictions are compared with the ground truth obtained by crowdsourcing tools. Results show that unusualness is the most useful cue to predict interestingness for strong context, whereas general preferences are more relevant for weaker contexts. These experiments highlight the importance of the image type, layout and context when estimating its interestingness.

Recently, crowdsourcing tools are employed to obtain less subjective and more reliable annotations of the datasets. However, it brings new problems such as sparse and outliers annotations. While above approaches [14, 15] prune the annotation outliers by majority voting, a new approach to globally detect outliers is proposed in [16]. They propose a Unified Robust Learning to Rank (URLR) framework to identify annotation outliers and, simultaneously, to build an interestingness ranking. This method is applied in WD [14, 15] and in a YouTube dataset [21] for image and video interestingness respectively (details of video results in Section 2.2). The experiments of image ranking prediction outperforms [15], proving the efficiency of the novel outlier detection model.

In contrast with [14–16], Dhar et al. [5] use the actual interestingness in social networks (number of views, popularity of user, etc.) to obtain the ground truth in order to avoid the crowdsourcing drawbacks and compare predictions with real behavior since, usually, the users concept of interestingness is not consistent with their behavior on social networks when selecting images to share or like. They study how aesthetic cues may be useful to predict image interestingness over Flickr photos, becoming a benchmark for most of the subsequent works on this research branch. The studied cues, referred as *high level describable attributes*, are: compositional attributes (layout of images), content attributes (presence of specific objects, categories of objects) and Sky-illumination attributes (natural outdoor illumination). They train a SVM to predict both image aesthetic and interestingness. Precision-Recall curves show good performance using high level attributes and better results when combining with low level features.

Some other approaches are focused on the prediction of image interestingness using social platforms information such as the number of views, likes or shares an image receives as a complement of visual cues. In [17], authors distinguish between Visual Interestingness (VI) and Social Interestingness (SI) since an image could be considered interesting due to its visual content or its social context. The VI score of each image is related to the crowdsourcing results whereas the SI score is provided by the photo sharing services Flickr and Pinterest and depends on statistics such as the number of likes, comments or number of users sharing the image. Hsieh et al. [17] investigate the correlation between VI, SI, and image aesthetics. Results show, more formally than [5, 15], a high correlation between VI and image aesthetics. It is also exposed the small or null correlation between VI and SI indicating that beautiful images are not the more likely to be shared by users in social networks. They also build a predictor using low level features such as color, edge, texture and saliency. Results show that texture and color are the best features to estimate VI and SI respectively, although the experiments only use low level features which does not allow the comparison with [5] that, in contrast with this work, obtains the best results with high level features.

In [18], they explore more in depth the social cues (amount of followers, number and content of tags, uploaded time, etc.) related to SI, referred as popularity, over Instagram. They also consider semantic concepts of images that refer to the objects depicted on the images and image categories. Experiments show: (i) the most correlated cue with popularity is the number of followers. This is not a surprising result since the more followers a user has the more people will view and share its images; (ii) low correlation between the image semantic concept and its popularity in the social network. They compare popularity on Instagram with crowdsourcing scores and results show that images depicting people, stadiums, baseball or amusement parks are the most popular on Instagram whereas the least popular contain buildings, fountains or cityscapes. In contrast, the user study selects images containing cityscapes, animals, restaurants or fountains as the most popular and people and buildings images as the least popular.

These conflicting results evidence the importance to consider SI to predict if an image will be interesting or popular in social networks. With this aim [19] investigates how image content and social context affect image popularity. They use features based on image content such as simple image features (hue, saturation and value), low level features (gist, texture, color patches, gradient and a learned interestingness score) and high level features (objects), whereas the features to highlight social cues are the number of views, total images uploaded by the user, number of contacts, etc. Used images are extracted from Flickr, which provides all the required social information. A SVR is trained to predict image interestingness using the aforementioned features. Results show social cues are better to estimate the number of views of an image and these are improved when social context features are complemented with image content features. They also build popularity maps of images to visualize which image regions are more influential and better understand image interestingness. Similar approaches investigate which parts of images humans look at [22] or which words are the most dominant to describe an image [23–25]. These methods can be used to generate textual descriptions of images or sort retrieved results, on image search, depending on the dominance of the attributes provided in the query.

In Table 1 we compare the best results of the studied method. We also detail their contribution and the attributes, training methods and datasets used.

2.2 What about video?

In addition to image, the issue of how to measure video interestingness on internet is addressed from the computer vision community. Due to the subjectivity of interestingness, many papers use social cues for video popularity prediction [29–32]. However, video content is also relevant and image and audio features can be used. Thus, from the computer vision perspective, some papers have been published [16, 21, 33] since the first approach in 2009 [34].

The method proposed in [34] leverages Flickr images to obtain the interestingness of each frame for YouTube videos with the key point that frames similar to interesting images should also be interesting. The similarity between images depends on the scene content (SIFT features) and composition (similar content in similar location). Although preliminary results are encouraging, the experiment is restricted to travel videos that contain well-known places and only the frames similar to these famous scenes are considered interesting. With these constraints, results may not be extendable to other less bounded experiments. Another drawback of this method is that the selected images from Flick need to be manually clustered depending on the depicted scene.

Actually, Liu et al. [34] only take into account the similarity between the video frames and interesting images, what is, in fact, the same as image interestingness prediction. On the other hand, Jiang et al. [21] proposes an entire video-level prediction using visual (SIFT, HOG, SSIM and GIST), audio (Mel-Frequency Cepstral Coefficients (MFCC), Spectrogram SIFT and six basic audio descriptors) and high level semantic features (object, scenes and photographic style). Authors propose a model to rank the videos instead of predicting an interestingness score. To evaluate this model they construct two benchmark datasets of Flickr and YouTube videos whose ground-truth is collected by crowdsourcing. Individual results show that videos are better ranked with visual and audio features, and their combination provides the best results. Otherwise, in contrast to [5] for image, high level attributes give the worst performance.

The approach for global outliers detection of the crowdsourcing annotations detailed in the previous section [18] also provides results for video interestingness prediction. Fu et al. apply their method (URLR) to rank videos from the dataset

Table 1: Comparison of image interestingness methods in 5 top rows of the table. Comparison of image memorability methods in 4 top rows of the table.

Authors	Contribution	Attributes	Training	Dataset[a]	Result
Dhar et al. 2011 [5]	- High level describable attributes - Predictor of aesthetics - Predictor of interestingness	Compositional (salient objects, rule of thirds, depth of field and opposing colors), content (presence of people or animals, portrait, indoor-outdoor and scene type) and sky-Illumination (clear, cloudy or sunset).	SVM	FD	Prec-Recall curves
Grabner et al. [14]	- Predictor of interestingness in image sequences	Emotion (depending on brightness an saturation), complexity (bytes of encoded image), novelty (outlier detection) and learned interestingness score.	ν-SVR	WD	0.36[c]
Gygli et al. 2013 [15]	- Predictor of interestingness in different image context.	Unusualness (global outliers and composition of parts), aesthetics (color, arousal, complexity, contrast and edge distribution) and general preferences.	ν-SVR	WD SCD MD	0.42[c] 0.83[c] 0.77[c]
Hsieh et al. 2014 [17]	- Comparison of visual and social interestingness and aesthetics. - Predictor of VI and SI separately.	- **Color** - Texture - Saliency - Edge	Adaboost	PD	0.73[bd]
Khosla et al. 2014 [19]	- Importance of image content and social cues for popularity. - Predictor of image popularity. - Visualization of popularity of image regions.	- Image content: simple image features (hue, saturation, value and intensity), low-level features (gist, texture, color patches, gradient and a learned score) and high-level features (objects). - Social cues: mean views, photo count, contacts, groups, group members, member duration, is pro, tags, title length and description length.	SVR	VSOD UMD USD	0.81[e] 0.72[e] 0.48[e]
Isola et al. 2011 [6]	- Analysis of relevant features for image memorability. - Predictor of image memorability. - Memorability map.	Simple image features (hue, saturation, value and intensity), non-semantic object statistics (object counts and areas), **semantic object statistics** (labeled object counts and areas), **scene category and global features** (GIST, SIFT, HOG2x2 and SSIM).	SVR	MD	0.54[be]
Isola et al. 2011 [26]	- More understandable features for image memorability. - Predictor of image memorability	- General attributes: spatial layout (enclosed /open, empty/cluttered), aesthetics (dull/ attractive, pleasant), emotions (Funny?, frightening?), actions, location and people. - People attributes: visibility, gender, age, hair length and color, clothing, activities, accessories, subject and scenario. - Global features and Object and Scene annotation from [6]	SVR	MD	0.55[e]
Khosla et al. 2012 [27]	- Memorability maps - Predictor of image memorability.	- Gradient - Color - Texture - Saliency - Shape - Semantic	SVM-Rank	MD	0.50[e]
Kim et al. 2013 [28]	- Two spatial features: Weighted Object Area (WOA) and Relative Area Rank (RAR)	WOA, RAR, global features and Scene annotations from [6] and attribute annotation from [26]	ϵ-SVR	MD	0.58[e]

[a] Databases are datailed in Section 4 [b] Best results uses only the attributes in bold
[c] Average precision value [d] Classification rate [e] Spearman's rank correlation (ρ)

built by Jiang et al. [21](YTD). Comparing both methods, URLD is superior and extends its good results, not only for image, but also for video prediction.

In [33] a different approach to measure video interestingness is proposed relating interestingness to how appealing or curious is a scene. With this insight, several features to highlight affectivity, aesthetics and semantic content are extracted to find appealing scenes in a video. They train two frameworks: a binary (positive/negative labels) and a graded (multiple degrees of annotations) relevance systems, the latter yields the best results when combining all the features.

In contrast with the above reviewed work, other approaches consider video interestingness as a subjective property and tackle the prediction issue using social information such as users patterns and current tendencies on video sharing services. A benchmark work is [29] that analyze the popularity distribution and evolution of videos from YouTube and Daum Videos (popular Korean platform). Moreover, duplicated and illegal content is posed as potential problems for accurate video popularity ranking. The growth patterns of video popularity are studied in [30] over three video datasets: (i) the top lists videos, (ii) removed videos due to copyright violation and (iii) random queries videos. In addition, the mechanisms to attract users towards a video are specified finding that the internal web systems are the most relevant. Later, Figueiredo [31] present a methodology to predict trends and hits in user generated videos and Pinto et al. [32] propose to predict video popularity using early view patterns.

2.3 Conclusions

Studied papers from state-of-the-art aim to better understand which visual features make an image interesting and build predictors to estimate if an image will be shared or liked in social platforms. Studied features evidence that high level features are more useful than low level features and the high importance of social cues to predict image interestingness, and especially, when results are compared with the users behavior on social networks. They also find a low correlation between social interestingness in the net and visual interestingness and aesthetics. Combining social accounts information with low and high level features gives the best predictions of the user's behavior in photo sharing platforms.

Research on video encourage the study of visual and audio features for video retrieval that appears to be, contrary to image, more useful than high level features. As for image, aesthetic features are related with video interestingness.

3 Memorability

Image memorability has been studied by psychologists since the 70s. L. Standing [35, 36] was one of the first to study the capacity of people memorizing images. Results evidenced the large memory of human being and posed the need to continue investigating. In the past 10 years, psychological researches of visual short- and long-term memory (VSTM and VLTM) have presented important results. The role of VSTM and VLTM in natural scenes perception and visual search is studiend in [37]. Other studies have focused on the amount and precision of details humans can remember in VLTM [38–40]. Contrary to the assumption that large amount of images can be remembered but with few details, results from [38] indicate that VLTM can store a large amount of image details.

Over the past few years, not only in the psychological domain, but also computer vision researchers have shown interest for the study of visual memory [6, 26, 28, 27, 41, 42]. Image memorability is considered an intrinsic property of images that do not depend on the observer and can be explained in terms of image features. Several recent works [8, 42, 43] aim to explain why some pictures are more memorable than others finding out the most relevant features to predict image memorability. Most authors analyze the whole image, whereas few of them are focused on analyzing which image regions are more memorable [27, 41].

3.1 State-of-the-art

To the best of our knowledge, the earliest paper is from 2011 [6] and is a benchmark for the subsequent works [26, 28, 27, 41]. Isola et al. [6] introduce the insight of memorability as an intrinsic property of images highly independent of the user context. This statement is supported by a user study to build a benchmark dataset (MD) that shows a high agreement between participants when memorizing images. To study which features are more significant, several are considered and they conclude that best results are obtained with object and scene semantics features, achieving the best results when combining both of them. For memorability prediction they train a SVR to map from features to memorability scores and the best prediction is performed when combining all global features (GIST [20], SIFT [43], HOG2x2 [44–46] and SSIM [47]) with object and scene semantics features. As expected, results support the use of high level features, like [5] for interestingness, since they contribute with more image information and are closer to the attributes used by human when evaluating an image.

As an extension of [6], a deeper study of relevant features is found in [26]. They complement the dataset MD from [6] with more attribute annotations referred to spatial, content and aesthetic image properties. An information-theoretic approach is used to select a set of non-redundant features and calibrate them by maximizing mutual information with memorability. Experiments with the extended annotations outperform [6] that uses only object and scene annotations. Furthermore, the information-theoretic approach provides the relevance of each feature showing that images of enclosed spaces containing people with visible faces are memorable while landscapes and peaceful images are not. Moreover, contrary to popular belief, unusualness and aesthetic attributes are not related to high image memorability. Finally, they also propose an automatic prediction using learned annotations that outperforms previous work in [6].

Kim et al. [28] consider that, although unusualness is not related to memorability [26], unusualness of the expected object size or location could be relevant and introduce two novel spatial features. One takes into account the size and location of objects supporting the hypothesis that central objects are more likely to be remembered. The second captures the unusualness of each object size depending on the object class coverage size. They obtain similar results to [6, 26], but the main advantage is that these two features together perform better than object statistics from [6] and they do not required a high level human annotation.

In contrast to previous approaches, the image memorability research can be addressed by estimating which image regions may be forgotten [27] modeling automatic memorability maps and combining both local and global features with no required human annotations. Results with gradient, color, texture, saliency, shape and semantic meaning outperform state-of-the-art results [6].

As a complete novel point of view, [41] introduces the idea of visual inception that refers to the possibility of modifying the memorability of images. Experiments describe how different images with similar structure can be identified as the same image. They also pose the issue of how an image can be changed while still making people believe they have seen it. However, the experiments are not enough detailed to show consistent results. A similar approach for face photos is [42] that manipulates traits of faces to make more memorable images.

In Table 1 we compare the best predictions of the studied method. We also detail their contribution and the attributes, training methods and datasets used.

3.2 Conclusions

The publications reviewed aim to understand and predict automatically image memorability. As for interestingness, features analysis reveal the relevance of high level features due to their high contribution of image information, and find images of enclosed spaces containing people with visible faces are memorable while landscapes and peaceful images are not. Moreover, contrary to popular belief, unusual and beautiful images are not necessarily memorable, unlike for interestingness where these attributes are influential. These assumptions are supported by memorability maps generated to visualize the memorable and forgettable regions. Good prediction results support the idea that memorability is a property independent of the user context and can be predicted in terms of image features.

4 Image and Video Datasets

We detail below the different datasets used for the experiments described above:

- **Flickr dataset (FD):** 40,000 images from Flickr. It is used in [5] for image interestingness prediction.
- **Webcam dataset (WD):** 20 sequences of 159 images each, from different public webcams. It is used in [14–16] for image interestingness prediction.
- **Scene categories dataset (SCD) [23]:** 2,688 images from 8 scene categories. It is used in [15] for image interestingness prediction.
- **Memorability dataset (MD):** 2,222 images. It is used in [6, 26, 28, 27] for memorability prediction and in [15] for image interestingness prediction.
- **Pinterest dataset (PD):** 989 images from Pinterest. It is used in [17] for image interestingness prediction.
- **Visual Sentiment Ontology dataset (VSOD):** 930k images collected from 400k users of Flickr. In [19], it is referred as one-per-user dataset and used to predict image interestingness.

- **User-mix dataset (UMD):** 1.4M images of a subset of user from VSOD. It is created by selecting 100 users from VSOD that have between 10k and 20k shared images. It is used in [19] to predict image interestingness.
- **User-specific dataset (USD):** independent datasets with the images of each user from UMD. It is used in [19] for image interestingness prediction.
- **YouTube Travel dataset (YTTD):** 3 videos for each of the 10 queries from YouTube travel category. It is used for video interestingness in [34].
- **Flickr Video dataset (FVD):** 400 videos for each search (15 queries). It is used in [21] for video interestingness prediction.
- **YouTube dataset (YTD):** 30 videos for each query (14 queries). It is used in [16, 21] for video interestingness prediction.

5 Conclusions and Future Work

A review of image interestingness and memorability research has been presented. Although these properties have been studied from psychologists since the 70s, they are new on the computer vision area, the first paper is from 2011. For image interestingness, experiments show a need to differentiate between visual and social interestingness since the interesting images collected by crowdsourcing are not the same than the interesting images in social platforms. They conclude that: (i) high level features are the most useful to predict visual interestingness; (ii) image aesthetics are highly correlated with it; (iii) for social interestingness, the social features are the most relevant and their combination with image features gives the best results. Similar conclusions can be found for memorability where high level features perform the best results. However, in contrast to interestingness, memorability has a low correlation with aesthetics and unusualness.

Some of these findings can be extended to video such as the correlation between interestingness and aesthetics and the effectiveness of social context based methods. In contrast with image, low level image and audio features work better than high level features. Future work on video analysis should address the interestingness prediction combining image and audio features with social cues.

Memorability is an objective property and image features provide good results, while interestingness is more subjective and depends largely on social context. Most works consider all users have the same context assuming the interestingness objetivity. Future research should be addressed to find common image features of most interesting images from a single user. Image features from the whole social network should also be studied to find similarities between images.

Further research on how to monetize the large amount of visual data uploaded in internet should be addressed. Interestingness and memorability scores will be used to appraise images, thus, companies will be provided with the most socially interesting and memorable images. Images owners that allow their use are rewarded with profitable assets. Also, the presence of brands could be considered to compute different image appraisal depending on each company.

Acknowledgments. This work was supported in part by VISUAL AD, RTC-2014-2566-7 and GEPID, RTC-2014-2576-7, as well as the consolidated research group CSI ref.2014 SGR 1469, and the grant of industrial PhD with expedient number 2014 DI 007.

References

1. Pang, B., Lee, L.: Opinion mining and sentiment analysis. Found. Trends Inf. Retr. **2**(1-2) (2008) 1–135
2. Bifet, A., Frank, E.: Sentiment knowledge discovery in twitter streaming data. In: Discovery Science, Springer (2010) 1–15
3. Liu, B., Hu, M., Cheng, J.: Opinion observer: analyzing and comparing opinions on the web. In: 14th int. conf. on World Wide Web, ACM (2005) 342–351
4. Mitrović, M., Paltoglou, G., Tadić, B.: Quantitative analysis of bloggers' collective behavior powered by emotions. J. Stat. Mech: Theory Exp. (02) (2011) P02005
5. Dhar, S., Ordonez, V., Berg, T.L.: High level describable attributes for predicting aesthetics and interestingness. In: CVPR, IEEE (2011) 1657–1664
6. Isola, P., Xiao, J., Torralba, A., Oliva, A.: What makes an image memorable? In: CVPR, IEEE (2011) 145–152
7. Silberschatz, A., Tuzhilin, A.: On subjective measures of interestingness in knowledge discovery. In: KDD. (1995) 275–281
8. Tan, P.N., Kumar, V., Srivastava, J.: Selecting the right interestingness measure for association patterns. In: ACM SIGKDD. (2002) 32–41
9. Schaul, T., Pape, L., Glasmachers, T., Graziano, V., Schmidhuber, J.: Coherence progress: a measure of interestingness based on fixed compressors. In: Artif. Gen. Intell. Springer (2011) 21–30
10. Berlyne, D.E.: Conflict, arousal, and curiosity. (1960)
11. Chen, A., Darst, P.W., Pangrazi, R.P.: An examination of situational interest and its sources. British Journal of Educational Psychology **71**(3) (2001) 383–400
12. Smith, C.A., Ellsworth, P.C.: Patterns of cognitive appraisal in emotion. Journal of personality and social psychology **48**(4) (1985) 813
13. Turner Jr, S.A., Silvia, P.J.: Must interesting things be pleasant? a test of competing appraisal structures. Emotion **6**(4) (2006) 670
14. Grabner, H., Nater, F., Druey, M., Van Gool, L.: Visual interestingness in image sequences. In: ACM Int. Conf. Multimed. (2013) 1017–1026
15. Gygli, M., Grabner, H., Riemenschneider, H., Nater, F., Gool, L.V.: The interestingness of images. In: ICCV, IEEE (2013) 1633–1640
16. Fu, Y., Hospedales, T.M., Xiang, T., Gong, S., Yao, Y.: Interestingness prediction by robust learning to rank. In: ECCV. Springer (2014) 488–503
17. Hsieh, L.C., Hsu, W.H., Wang, H.C.: Investigating and predicting social and visual image interestingness on social media by crowdsourcing, In: ICASSP IEEE(2014)
18. Fiolet, E.: Analyzing image popularity
19. Khosla, A., Das Sarma, A., Hamid, R.: What makes an image popular? In: World Wide Web. (2014) 867–876
20. Oliva, A., Torralba, A.: Modeling the shape of the scene: A holistic representation of the spatial envelope. Int. J. Comput. Vis. **42**(3) (2001) 145–175
21. Jiang, Y.G., Wang, Y., Feng, R., Xue, X., Zheng, Y., Yang, H.: Understanding and predicting interestingness of videos. In: AAAI. (2013)
22. Judd, T., Ehinger, K., Durand, F., Torralba, A.: Learning to predict where humans look. In: Computer Vision, IEEE (2009) 2106–2113
23. Spain, M., Perona, P.: Measuring and predicting importance of objects in our visual world. Tech. Rep. (2007)

24. Berg, A.C., Berg, T.L., Daume, H., Dodge, J., Goyal, A., Han, X., et al.: Understanding and predicting importance in images. In: CVPR, IEEE (2012) 3562–3569
25. Turakhia, N., Parikh, D.: Attribute dominance: What pops out? In: ICCV, IEEE (2013) 1225–1232
26. Isola, P., Parikh, D., Torralba, A., Oliva, A.: Understanding the intrinsic memorability of images. In: NIPS. (2011) 2429–2437
27. Khosla, A., Xiao, J., Torralba, A., Oliva, A.: Memorability of image regions. In: Advances in Neural Information Processing Systems. (2012) 305–313
28. Kim, J., Yoon, S., Pavlovic, V.: Relative spatial features for image memorability. In: ACM Multimedia. (2013) 761–764
29. Cha, M., Kwak, H., Rodriguez, P., Ahn, Y.Y., Moon, S.: Analyzing the video popularity characteristics of large-scale user generated content systems. IEEE/ACM Transactions on Networking (TON) **17**(5) (2009) 1357–1370
30. Figueiredo, F., Benevenuto, F., Almeida, J.M.: The tube over time: characterizing popularity growth of youtube videos. In: ACM WSDM. (2011) 745–754
31. Figueiredo, F.: On the prediction of popularity of trends and hits for user generated videos. In: ACM WSDM. (2013) 741–746
32. Pinto, H., Almeida, J.M., Gonçalves, M.A.: Using early view patterns to predict the popularity of youtube videos. In: ACM WSDM. (2013) 365–374
33. Redi, M., Merialdo, B.: Where is the beauty?: Retrieving appealing videoscenes by learning flickr-based graded judgments. In: ACM Multimedia. (2012) 1363–1364
34. Liu, F., Niu, Y., Gleicher, M.: Using web photos for measuring video frame interestingness. In: IJCAI. (2009) 2058–2063
35. Standing, L., Conezio, J., Haber, R.N.: Perception and memory for pictures: Single-trial learning of 2500 visual stimuli. Psychonomic Science **19**(2) (1970) 73–74
36. Standing, L.: Learning 10000 pictures. Q. J. Exp. Psychol. **25**(2) (1973) 207–222
37. Hollingworth, A.: Constructing visual representations of natural scenes: the roles of short-and long-term visual memory. J. Exp. Psychol.: Hum. Percept. Perform. **30**(3) (2004) 519
38. Brady, T.F., Konkle, T., Alvarez, G.A., Oliva, A.: Visual long-term memory has a massive storage capacity for object details. Nat. Acad. Sci. (2008) 14325–14329
39. Konkle, T., Brady, T.F., Alvarez, G.A., Oliva, A.: Scene memory is more detailed than you think the role of categories in visual long-term memory. Psychological Science **21**(11) (2010) 1551–1556
40. Konkle, T., Brady, T.F., Alvarez, G.A., Oliva, A.: Conceptual distinctiveness supports detailed visual long-term memory for real-world objects. Journal of Experimental Psychology: General **139**(3) (2010) 558
41. Khosla, A., Xiao, J., Isola, P., Torralba, A., Oliva, A.: Image memorability and visual inception. In: SIGGRAPH Asia 2012 Technical Briefs, ACM (2012) 35
42. Khosla, A., Bainbridge, W.A., Torralba, A., Oliva, A.: Modifying the memorability of face photographs. In: ICCV, IEEE (2013) 3200–3207
43. Lazebnik, S., Schmid, C., Ponce, J.: Beyond bags of features:spatial pyramid matching for recognizing natural scene categories, CVPR IEEE(2006) 2169–2178
44. Felzenszwalb, P.F., Girshick, R.B., McAllester, D., Ramanan, D.: Object detection with discriminatively trained part-based models. IEEE PAMI (2010) 1627–1645
45. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: CVPR. Volume 1., IEEE (2005) 886–893
46. Xiao, J., Hays, J., Ehinger, K.A., Oliva, A., Torralba, A.: Sun database: Large-scale scene recognition from abbey to zoo. In: CVPR, IEEE (2010) 3485–3492
47. Shechtman, E., Irani, M.: Matching local self-similarities across images and videos. In: CVPR, IEEE (2007) 1–8