Sentribute: Image Sentiment Analysis from a Mid-level Perspective

Jianbo Yuan
Department of Electrical & Computer Engineering
University of Rochester
Rochester, NY 14627
jyuan10@ece.rochester.edu

Quanzeng You
Department of Computer Science
University of Rochester
Rochester, NY 14627
qyou@cs.rochester.edu

Sean Mcdonough
Department of Electrical & Computer Engineering
University of Rochester
Rochester, NY 14627
smcdono2@u.rochester.edu

Jiebo Luo
Department of Computer Science
University of Rochester
Rochester, NY 14627
jluo@cs.rochester.edu

ABSTRACT

Visual content analysis has always been important yet challenging. Thanks to the popularity of social networks, images become an convenient carrier for information diffusion among online users. To understand the diffusion patterns and different aspects of the social images, we need to interpret the images first. Similar to textual content, images also carry different levels of sentiment to their viewers. However, different from text, where sentiment analysis can use easily accessible semantic and context information, how to extract and interpret the sentiment of an image remains quite challenging. In this paper, we propose an image sentiment prediction framework, which leverages the mid-level attributes of an image to predict its sentiment. This makes the sentiment classification results more interpretable than directly using the low-level features of an image. To obtain a better performance on images containing faces, we introduce eigenface-based facial expression detection as an additional mid-level attributes. An empirical study of the proposed framework shows improved performance in terms of prediction accuracy. More importantly, by inspecting the prediction results, we are able to discover interesting relationships between mid-level attribute and image sentiment.

Categories and Subject Descriptors

H.2.8 [**Database management**]: Database Applications; H.3.1 [**Information Storage and Retrieval**]: Content Analysis and Retrieval; I.5.4 [**Pattern Recognition**]: Applications

WISDOM '13, August 11 2013, Chicago, USA

General Terms

Algorithms, Experimentation, Application

Keywords

Image sentiment, Analysis, Mid-level Attributes, Visual Content

1. INTRODUCTION

Nowadays, social networks such as Twitter and microblog such as Weibo become major platforms of information exchange and communication between users, between which the common information carrier is tweets. A recent study shows that images constitute about 36 percent of all the shared links on Twitter[1], which makes visual data mining an interesting and active area to explore. As an old saying has it, an image is worth a thousand words. Much alike textual content based mining approach, extensive studies have been done regarding aesthetics and emotions in images [3, 8, 28]. In this paper, we are focusing on sentiment analysis based on visual information analysis.

So far analysis of textual information has been well developed in areas including opinion mining [18, 20], human decision making [20], brand monitoring [9], stock market prediction [1], political voting forecasts [18, 25] and intelligence gathering [31]. Figure 1 shows and example of image tweets. In contrast, analysis of visual information covers areas such as image information retrieval [4, 33], aesthetics grading [15] and the progress is relatively behind.

Social networks such as Twitter and microblogs such as Weibo provide billions of pieces of both textual and visual information, making it possible to detect sentiment indicated by both textual and visual data respectively. However, sentiment analysis based on a visual perspective is still in its infancy. With respect to sentiment analysis, much work has been done on textual information based sentiment analysis [18, 20, 29], as well as online sentiment dictionary [5, 24].

[1]http://socialtimes.com/is-the-status-update-dead-36-of-tweets-are-photos-infographic_b103245#.UDLhTK9rHY8.wordpress

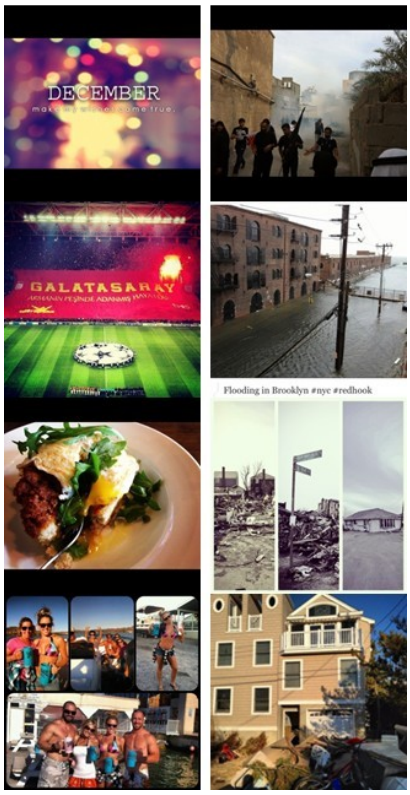

Figure 1: Selected images crawled from Twitter showing (left column) positive sentiment and (right column) negative sentiments.

Semantics and concept learning approaches [6, 19, 16, 22] based on visual features is another way of sentiment analysis without employing textual information. However, semantics and concept learning approaches are hampered by the limitations of object classifier accuracy The analysis of aesthetics [3, 15], interestingness [8] and affect or emotions [10, 14, 17, 32] of images are most related to sentiment analysis based on visual content. Aiming to conduct visual content based sentiment analysis, current approaches includes employing low-level features [10, 11, 12], via facial expression detection [27] and user intent [7]. Sentiment analysis approaches based on low-level features has the limitation of low interpretability, which in turn makes it undesirable for high-level use. Metadata of images is another source of information for high-level feature learning [2]. However, not all images contain such kind of data. Therefore, we proposed Sentribute, an image sentiment analysis algorithm based on mid-level features.

Compared to the state-of-the-art algorithms, our main contribution to this area is two-fold: first, we propose Sentribute, an image-sentiment analysis algorithm based on 102 mid-level attributes, of which results are easier to interpret and ready-to-use for high-level understanding. Second, we introduce eigenface to facial sentiment recognition as a solution for sentiment analysis on images containing people. This is simple but powerful, especially in cases of extreme facial expressions, and contributed an 18% gain in accuracy over decision making only based on mid-level attributes, and 30% over the state of art methods based on low level features.

The remainder of this paper is organized as follows: in Section 2, we present an overview of our proposed Sentribute framework. Section 3 provides details for Sentribute, including low-level feature extraction, mid-level attribute generation, image sentiment prediction, and decision correction based on facial sentiment recognition. Then in Section 4, we test our algorithm on 810 images crawled from Twitter and make a comparison with the state of the art method, which makes prediction based on low-level features and textual information only. Finally, we summarize our findings and possible future extensions of our current work in Section 5.

2. FRAMEWORK OVERVIEW

Figure 2 presents our proposed Sentribute framework. The idea for this algorithm is as follows: first of all, we extract scene descriptor low-level features from the SUN Database [7] and use these four features to train our classifiers by Liblinear [10] for generating 102 predefined mid-level attributes, and then use these attributes to predict sentiments. Meanwhile, facial sentiments are predicted using eigenfaces. This method generates really good results especially in cases of predicting strong positive and negative sentiments, which makes it possible to combine these two predictions and generate a better result for predicting image sentiments with faces. To illustrate how facial sentiment help refine our prediction based on only mid-level attributes, we present an example in Section 4, of how to correct our false positive/negative prediction based on facial sentiment recognition.

3. SENTRIBUTE

In this section we outline the design and construction of the proposed Sentribute, a novel image sentiment prediction method based on mid-level attributes, together with a decision refine mechanism for images containing people. For image sentiment analysis, we conclude the procedure starting from dataset introduction, low-level feature selection, building mid-level attribute classifier, image sentiment prediction. As for facial sentiment recognition, we introduce eigenface to fulfill our intention.

3.1 Dataset

Our proposed algorithm mainly contains three steps: first is to generate mid-level attributes labels. For this part, we train our classifier using SUN Database[2], the first large-scale scene attribute database, initially designed for high-level scene understanding and fine-grained scene recognition [21]. This database includes more than 800 categories and 14,340 images, as well as discriminative attributes labeled by crowd-sourced human studies. Attributes labels are presented in form of zero to three votes, of which 0 vote means this image is the least correlated with this attribute, and three votes means the most correlated. Due to this voting mechanism, we have an option of selecting which set of images to be labeled as positive: images with more than one vote, introduced as soft decision (SD), or images with more than two votes, introduced as hard decision (HD).

Second step of our algorithm is to train sentiment predicting classifiers with images crawled from Twitter together

[2] http://groups.csail.mit.edu/vision/SUN/

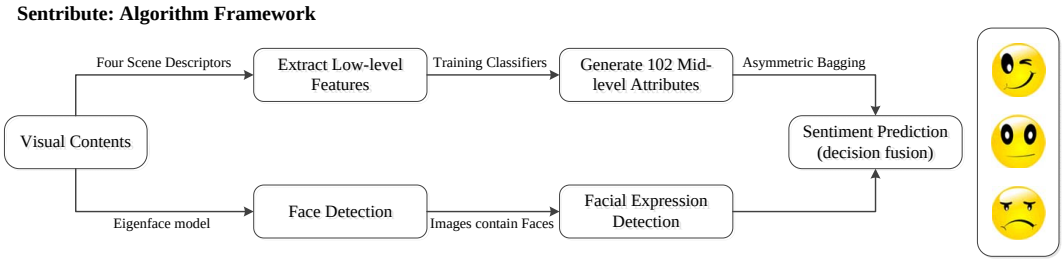

Figure 2: Selected images crawled from Twitter showing (a) positive sentiment and (b) negative sentiments.

with their textual data covering more than 800 images. Twitter is currently one of the most popular microblog platforms. Sentiment ground truth is obtained from visual sentiment ontology[3] with permission of the authors. The dataset includes 1340 positive, 223 negative and 552 neutral image tweets. For testing, we randomly select 810 images, only containing positive (660 tweets) and negative (150 tweets). Figure 1 shows images chosen from our dataset as well as their sentiment labels.

The final step is facial emotion detection for decision fusion mechanism. We chose to use the Karolinska Directed Emotional Faces dataset [13] mainly because the faces are all well aligned with each other and have consistent lighting, which makes generating good eigenface much easier. The dataset contains 70 men and women over two days expressing 7 emotions (scared, anger, disgust, happy, neutral, sad, and surprised) in five different poses (front, left prole, right prole, left angle, right angle).

3.2 Feature Selection

In this section, we are aiming to select low-level features for generating mid-level attributes, and we choose four general scene descriptor: gist descriptor [17], HOG 2x2, self-similarity, and geometric context color histogram features [30]. These four features were chosen because they are each individually powerful and because they can describe distinct visual phenomena in a scene perspective other than using specific object classifier. These scene descriptor features suffer neither from the inconsistent performance compared to commonly used object detectors for high-level semantics analysis of an image, nor from the difficulty of result interpretation generated based on low-level features.

3.3 Generating Mid-level Attribute

Given selected low-level features, we are then able to train our mid-level attribute classifiers based on SUN Database. We have 14,340 dimensions of sampling space, and over 170,000 dimensions of feature space. For classifier options, Liblinear[4] outperforms against LibSVM[5] in cases where the number of samples are huge and the number of feature dimension is huge. Therefore we choose Liblinear toolbox to implement SVM algorithm to achieve time saving.

The selection of mid-level attribute also plays an important part in image sentiment analysis. We choose 102 predefined mid-level attributes based on the following criteria: (1) have descent detection accuracy, (2) potentially correlated to one sentiment label, and (3) easy to interpret. We then select four types of mid-level attributes accordingly: (1) Material: such as metal, vegetation; (2) Function: playing, cooking; (3) Surface property: rusty, glossy; and (4) Spatial Envelope [17]: natural, man-made, enclosed.

We conduct mutual information analysis to discover mid-level attributes that are most correlated with sentiments. For each mid-level attribute, we computed the MI value with respect to both positive and negative sentiment category (Figure 4). Table 1 illustrates 10 most distinguishable mid-level attributes for predicting both positive and negative labels in a descending order based on both SD and HD. Figure 6 demonstrates Average Precision (AP) for the 102 attributes we selected, for both SD and HD. It's not surprising to see that attributes of material (flowers, trees, ice, still water), function (hiking, gaming, competing) and spatial envelop (natural light, congregating, aged/worn) all play an important role according to the result of mutual information analysis

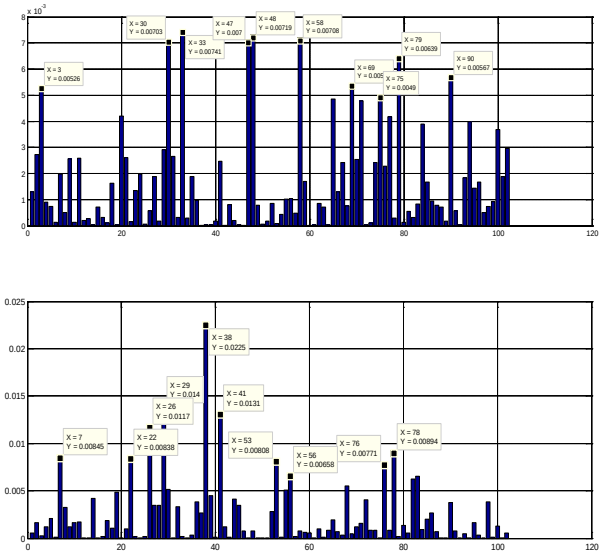

Figure 4: Computing Mutual Information for each label.

3.4 Image Sentiment Prediction

In our dataset we have 660 positive samples and 140 negative samples. It is likely to obtain a biased classifier based on these samples alone. Therefore we introduce asymmetric bagging [23] to dealing with biased dataset. Figure 6 presents the idea of asymmetric bagging: instead of build-

[3] http://visual-sentiment-ontology.appspot.com/
[4] http://www.csie.ntu.edu.tw/~cjlin/liblinear/
[5] http://www.csie.ntu.edu.tw/~cjlin/libsvm/

Figure 3: The images in the table above are grouped by the number of positive labels (votes) received from AMT workers. From left to right the visual presence of each attribute increases [21].

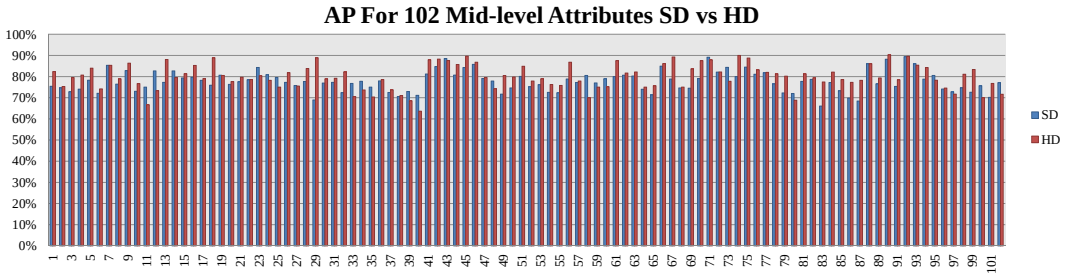

Figure 5: AP of the 102 Attributes based on SD and HD.

Table 1: Attributes with Top 10 Mutual Information.

TOP 10	Soft Decision	Hard Decision
1	congregating	railing
2	flowers	hiking
3	aged/worn	gaming
4	vinyl/linoleum	competing
5	still water	trees
6	natural light	metal
7	glossy	tiles
8	open area	direct sun/sunny
9	glass	aged/worn
10	ice	constructing

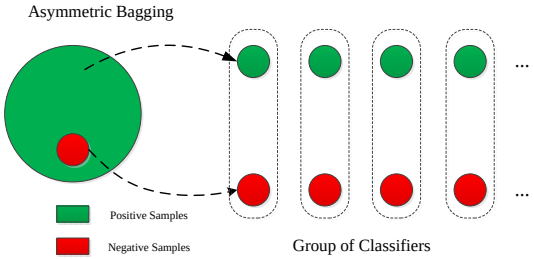

Figure 6: Asymmetric bagging.

ing one classifier, we now build several classifiers, and train them with the same negative samples together with different sampled positive samples of the same amount. Then we can combine their results and build an overall unbiased classifier.

3.5 Facial Sentiment Recognition

Our proposed algorithm, Sentribute, contains a final step of decision fusion mechanism by incorporating eigenface-based emotion detection approach. Images containing faces contribute to a great partition of the whole images that, 382 images from our dataset have faces. Therefore, facial emotion detection is not only useful but important for the overall performance of our algorithm.

In order to recognize emotions from faces we use classes of eigenfaces corresponding to different emotions. Eigenface was one of the earliest successful implementations of facial detection [26]; we modify the algorithm to be suitable for detecting classes of emotions. Though this method is widely appreciated already, we are the first to modify the algorithm to be suitable for detecting classes of emotions, and this method is simple yet surprisingly powerful for detecting facial emotions for front and consistent lightened faces. Note that we are not trying to propose an algorithm that

outperforms the state-of-the-art facial emotion detection algorithms. This is beyond the scope of this paper.

There are seven principal emotions that human's experience: scared (afraid), anger, disgust, happy, neutral, sad, and surprised. Due to the accuracy of the model and the framework of integrating the results with Sentribute, we reduce the set of emotions to positive, neutral, and negative emotions. This is done by classifying the image as one of the seven emotions and then mapping the happy and surprised emotions to positive sentiment, neutral sentiment to itself, and all other emotions to negative sentiment. At a high level, we are computing the eigenfaces for each class of emotion; we then compare the features of these eigenfaces to the features of the target image projected onto the emotion class space.

The algorithm requires a set of faces to train the classifier (more specifically to find the features of the images). We chose to use the Karolinska Directed Emotional Faces dataset [13] for many reasons, specifically the faces are all well aligned with each other and have consistent lighting, which makes generating good eigenfaces much easier. The dataset contains 70 men and women over two days expressing 7 emotions (scared, anger, disgust, happy, neutral, sad, and surprised) in five different poses (front, left profile, right profile, left angle, right angle). We use a subset of the KDEF database for our training set, only using the 7 frontal emotions from one photographing session.

Training the dataset and extracting the eigenfaces from the images of each emotion class was accomplished by using principal component extraction. We preprocess the training data by running it through fdlibmex[6], a fast facial detection algorithm to obtain the position and size of the face. We then extract the face from the general image and scale it to a 64×64 grayscale array; it is then vectored into a 4096 length vector. We concatenate the individual faces from each class into a $M \times N$ array $\mathbf{X}$, where M is the length of each individual image and N is the number of images in the class. We then are able to find the eigenfaces by using Principal Component Extraction. Principal component extraction converts correlated variables, in our case a set of images, into an uncorrelated variables via an orthogonal transform. We implement principal component analysis by first computing the covariance matrix

$$C = (x - \mu)(x - \mu)^T, \tag{1}$$

where μ is the mean of which has been concatenated to the same size of $\mathbf{X}$. The eigenvectors of C are then calculated are arranged by decreasing eigenvalues. Only the twenty largest eigenvectors are chosen for each class of facial emotions. The principle eigenfaces are simply the eigenvectors of the system that have the largest eigenvalues. We compute the features of the class as shown below.

$$\begin{aligned} E^C &= PCA(X^C) \\ F^C &= E^C(X^C - \mu^C) \end{aligned} \quad . \tag{2}$$

In order to classify the target image preprocessing is necessary to preprocess the image as we preprocess the training dataset, which we will denote y. The classification of a test face is performed by comparing the distance of the features of the target face (projected onto the emotion subspace) to the features of the eigenfaces of the subspace. We then choose the class that minimizes this function as the predicted class, specifically

$$\arg\min_C \sum_i \parallel E_i^C(y - \mu^C) - F_i^C \parallel, \tag{3}$$

where i is each individual feature column vector in the array [26].

We then set a threshold value, which was determined empirically, in order to filter out results that are weakly classified. In this case, no result is given. Figure 7 shows examples of classified facial emotions.

(a) Classification: Negative (b) Classification: Positive (c) Classification: Positive

Figure 7: Examples of Eigenface-based Emotion Detection.

4. EXPERIMENTS

4.1 Image Sentiment Prediction

As mentioned before, state-of-the-art sentiment analysis approach can be mainly concluded as: (1) textual information based sentiment analysis, as well as online sentiment dictionary [5, 24] and (2) sentiment analysis based on low-level features. Therefore, in this section, we set three baselines: (1) low-level feature based approach and (2) textual content based approach [24] and (3) online sentiment dictionary SentiStrength [5].

4.1.1 Image Sentiment Classification Performance

First we demonstrate results of our proposed algorithm, image sentiment prediction based on 102 mid-level attributes (SD vs. HD). Both Linear SVM and Logistic Regression algorithms are employed for comparison.

As demonstrated in Table 2, performance of precision for both Linear SVM and Logistic Regression outperforms over that of recall. Owing to the implementation of asymmetric bagging, we are now able to classify negative samples in a decent detection rate. Smaller number of false positive samples and relatively larger number of detected true positive samples contribute to this unbalanced value of precision and recall performance.

Table 2: Image Sentiment Prediction Performance.

		Precision	Recall	Accuracy
Linear SVM	SD	82.6%	56.8%	55.2%
	HD	86.7%	59.1%	61.4%
Logistic Regr	SD	84.3%	54.7%	54.8%
	HD	88.1%	58.8%	61.2%

The next thing we are interested in is the comparison against baseline algorithms.

[6] `http://www.mathworks.com/matlabcentral/fileexchange/20976`

4.1.2 Low-level Feature Based and Textual Content Based Baselines

For low-level feature based algorithm, Ji et al. employed the following visual features: a dimensional Color Histogram extracted from the RGB color space, a 512 dimensional GIST descriptor [17], a 53 dimensional Local Binary Pattern (LBP), a Bag-of-Words quantized descriptor using a 1000 word dictionary with a 2-layer spatial pyramid, and a 2659 dimensional Classemes descriptor. Both Linear SVM and Logistic Regression algorithms are used for classification. For textual content based algorithm, we choose Contextual Polarity, a phrase level sentiment analysis system [29], as well as SentiStrength API[7]. Table 3 the results of accuracy based on low-level features, mid-level attributes and textual contents.

Table 3: Accuracy of Sentiment Prediction.

(a) Comparison between low-level based algorithm and mid-level based algorithm.

	SVM (low)	Logistic Regr (low)	SVM (mid)
AC	50%	53%	61.4%

(b) Comparison between mid-level visual content based algorithm and textual content based algorithm.

	Contextual Polarity	SentiStrength	SVM (mid)
AC	61.7%	61%	61.4%

4.2 Decision Fusion

The final step of Sentribute is decision fusion. By applying eigenface-based emotion detection, we are able to improve the performance of our decision based on mid-level attributes only. We only take into account images with complete face with reasonable lighting condition. Therefore among all the images with faces, we first employ a face detection process and generate a set of 153 images as the testing data set for facial emotion detection and decision fusion. For each face we detected, we assigned them a label indicating sentiments: 1 for positive, 0 for neutral and -1 for negative sentiments. We thus computed a sentiment score for each image as a whole. For instance, if we detect three faces from an image, two of them are detected as positive and one of them is detected as neutral, then the overall facial sentiment score of this image is 2. These sentiment scores can be used for decision fusion with the decision made based on mid-level attributes only, i.e., we add up the facial sentiment score and the confidence score of the results based on mid-level attributes only returned by our classifiers to implement a decision fusion mechanism. Table 4 shows the improvements in accuracy after decision fusion.

Figure 8 presents examples of TP, FP, TN, FN samples generated by Sentribute. False classified samples show that it's hard to distinguish images only containing texts from both positive and negative labels, and images of big event / cerebration (football game or a concert) from those of protest demonstration. They both share similar general scene descriptors, similar lighting condition, and similar color tone. Another interesting false detected sample is the first image shown in false negative samples. Figures make frown expression on their faces, however the sentiment behind this expression is positive since they were meant to be funny. This sample is initially classified as positive based on mid-level attributes only, and then refined as negative because two strong negative facial expression are detected by our eigenface expression detector. This kind of images shows a better decision fusion metric would be one of our potential improvements.

[7] `http://sentistrength.wlv.ac.uk/`

Table 4: Accuracy of Sentribute Algorithm.

	Accuracy
Mid-level Based Prediction	64.71%
Facial Emotion Detection	73.86%
Sentribute (After Synthesis)	82.35%

5. CONCLUSION

In this paper we have demonstrated Sentribute, a novel image sentiment prediction algorithm based on mid-level attributes. Asymmetric bagging approach is employed to deal with unbalanced dataset. To enhance our prediction performance, we introduce eigenface-based emotion detection algorithm, which is simple but powerful especially in cases of detecting extreme facial expressions, to dealing with images containing faces and obtain a distinct gain in accuracy over result based on mid-level attributes only. Our proposed algorithm explores current visual content based sentiment analysis approach by employing mid-level attributes and without using textual content. We are aware that this work is just one out of many steps that several potential directions are exciting to set foot on. First, this mid-level based visual content can be introduced to aesthetics analysis as well. Also, a combination of our approach and textual content sentiment analysis approach might be beneficial. Additionally, further application of our proposed work includes but not limited to psychology, public opinion analysis and online activity emotion detection.

Acknowledgments

We thank Professor Shih-Fu Chang's group for providing us with the Columbia University data set for image sentiment analysis.

6. REFERENCES

[1] J. Bollen, H. Mao, and X. Zeng. Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1):1–8, 2011.

[2] E. Cambria and A. Hussain. Sentic album: content-, concept-, and context-based online personal photo management system. *Cognitive Computation*, 4(4):477–496, 2012.

[3] R. Datta, D. Joshi, J. Li, and J. Z. Wang. Studying aesthetics in photographic images using a computational approach. In *Computer Vision–ECCV 2006*, pages 288–301. Springer, 2006.

[4] R. Datta, D. Joshi, J. Li, and J. Z. Wang. Image retrieval: Ideas, influences, and trends of the new age. *ACM Computing Surveys (CSUR)*, 40(2):5, 2008.

[5] A. Esuli and F. Sebastiani. Sentiwordnet: A publicly available lexical resource for opinion mining. In *Proceedings of LREC*, volume 6, pages 417–422, 2006.

(a) True positive samples

(b) True negative samples

(c) False positive samples

(d) False negative samples

Figure 8: Examples of Sentiment Detection Results By Sentribute.

[6] A. Farhadi, I. Endres, D. Hoiem, and D. Forsyth. Describing objects by their attributes. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 1778–1785. IEEE, 2009.

[7] A. Hanjalic, C. Kofler, and M. Larson. Intent and its discontents: the user at the wheel of the online video search engine. In *Proceedings of the 20th ACM international conference on Multimedia*, pages 1239–1248. ACM, 2012.

[8] P. Isola, J. Xiao, A. Torralba, and A. Oliva. What makes an image memorable? In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 145–152. IEEE, 2011.

[9] B. J. Jansen, M. Zhang, K. Sobel, and A. Chowdury. Twitter power: Tweets as electronic word of mouth. *Journal of the American society for information science and technology*, 60(11):2169–2188, 2009.

[10] J. Jia, S. Wu, X. Wang, P. Hu, L. Cai, and J. Tang. Can we understand van gogh's mood?: learning to infer affects from images in social networks. In *Proceedings of the 20th ACM international conference on Multimedia*, pages 857–860. ACM, 2012.

[11] P. J. Lang, M. M. Bradley, and B. N. Cuthbert. International affective picture system (iaps): Technical manual and affective ratings, 1999.

[12] B. Li, S. Feng, W. Xiong, and W. Hu. Scaring or pleasing: exploit emotional impact of an image. In *Proceedings of the 20th ACM international conference on Multimedia*, pages 1365–1366. ACM, 2012.

[13] D. Lundqvist, A. Flykt, and A. Öhman. The karolinska directed emotional faces-kdef. cd-rom from department of clinical neuroscience, psychology section, karolinska institutet, stockholm, sweden. Technical report, ISBN 91-630-7164-9, 1998.

[14] J. Machajdik and A. Hanbury. Affective image classification using features inspired by psychology

and art theory. In *Proceedings of the international conference on Multimedia*, pages 83–92. ACM, 2010.

[15] L. Marchesotti, F. Perronnin, D. Larlus, and G. Csurka. Assessing the aesthetic quality of photographs using generic image descriptors. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 1784–1791. IEEE, 2011.

[16] M. R. Naphade, C.-Y. Lin, J. R. Smith, B. Tseng, and S. Basu. Learning to annotate video databases. In *SPIE Conference on Storage and Retrieval on Media databases*, 2002.

[17] A. Oliva and A. Torralba. Modeling the shape of the scene: A holistic representation of the spatial envelope. *International journal of computer vision*, 42(3):145–175, 2001.

[18] B. O'onnor, R. Balasubramanyan, B. R. Routledge, and N. A. Smith. From tweets to polls: Linking text sentiment to public opinion time series. In *Proceedings of the International AAAI Conference on Weblogs and Social Media*, pages 122–129, 2010.

[19] V. Ordonez, G. Kulkarni, and T. L. Berg. Im2text: Describing images using 1 million captioned photographs. In *Neural Information Processing Systems (NIPS)*, 2011.

[20] B. Pang and L. Lee. Opinion mining and sentiment analysis. *Foundations and trends in information retrieval*, 2(1-2):1–135, 2008.

[21] G. Patterson and J. Hays. Sun attribute database: Discovering, annotating, and recognizing scene attributes. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 2751–2758. IEEE, 2012.

[22] C. G. Snoek and M. Worring. Concept-based video retrieval. *Foundations and Trends in Information Retrieval*, 2(4):215–322, 2008.

[23] D. Tao, X. Tang, X. Li, and X. Wu. Asymmetric bagging and random subspace for support vector machines-based relevance feedback in image retrieval. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 28(7):1088–1099, 2006.

[24] M. Thelwall, K. Buckley, G. Paltoglou, D. Cai, and A. Kappas. Sentiment strength detection in short informal text. *Journal of the American Society for Information Science and Technology*, 61(12):2544–2558, 2010.

[25] A. Tumasjan, T. O. Sprenger, P. G. Sandner, and I. M. Welpe. Predicting elections with twitter: What 140 characters reveal about political sentiment. In *Proceedings of the fourth international AAAI conference on weblogs and social media*, pages 178–185, 2010.

[26] M. A. Turk and A. P. Pentland. Face recognition using eigenfaces. In *Computer Vision and Pattern Recognition, 1991. Proceedings CVPR'91., IEEE Computer Society Conference on*, pages 586–591. IEEE, 1991.

[27] V. Vonikakis and S. Winkler. Emotion-based sequence of family photos. In *Proceedings of the 20th ACM international conference on Multimedia*, MM '12, pages 1371–1372, New York, NY, USA, 2012. ACM.

[28] W. Wang and Q. He. A survey on emotional semantic image retrieval. In *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on*, pages 117–120. IEEE, 2008.

[29] T. Wilson, J. Wiebe, and P. Hoffmann. Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, pages 347–354. Association for Computational Linguistics, 2005.

[30] J. Xiao, J. Hays, K. A. Ehinger, A. Oliva, and A. Torralba. Sun database: Large-scale scene recognition from abbey to zoo. In *Computer vision and pattern recognition (CVPR), 2010 IEEE conference on*, pages 3485–3492. IEEE, 2010.

[31] C. C. Yang and T. D. Ng. Terrorism and crime related weblog social network: Link, content analysis and information visualization. In *Intelligence and Security Informatics, 2007 IEEE*, pages 55–58. IEEE, 2007.

[32] V. Yanulevskaya, J. Uijlings, E. Bruni, A. Sartori, E. Zamboni, F. Bacci, D. Melcher, and N. Sebe. In the eye of the beholder: employing statistical analysis and eye tracking for analyzing abstract paintings. In *Proceedings of the 20th ACM international conference on Multimedia*, MM '12, pages 349–358, New York, NY, USA, 2012. ACM.

[33] S. Zhang, M. Yang, T. Cour, K. Yu, and D. N. Metaxas. Query specific fusion for image retrieval. In *Computer Vision–ECCV 2012*, pages 660–673. Springer, 2012.