J Multimodal User Interfaces (2016) 10:95–97
DOI 10.1007/s12193-016-0213-z

EDITORIAL

Emotion recognition in the wild

Abhinav Dhall[1] · Roland Goecke[1] · Tom Gedeon[2] · Nicu Sebe[3]

Received: 2 February 2016 / Accepted: 4 February 2016 / Published online: 3 March 2016

For inferring the affective state of a person from data, captured in real-world conditions, methods which can perform emotion analysis 'in the wild' are required. Here, the term 'in the wild' signifies different environments/scenes and background noise, illumination conditions, head pose and occlusion. Automatic emotion recognition has made a significant progress in last two decades. However, such developed frameworks have been strictly employed to data collected in controlled laboratory settings with frontal faces, perfect illumination and posed expressions. On the contrary, images and videos on the WWW have been captured in different, unconstrained environments and this poses a big challenge to automatic facial emotion recognition methods. This special issue addresses the problem of emotion recognition in challenging conditions and is based on the recent series of Emotion recognition in the Wild (EmotiW) challenge.

Recently, the first EmotiW challenge [2] brought together researchers working on emotion recognition and the acted facial expressions in the wild (AFEW) [4] database formed the baseline for the challenge. AFEW has been created from movies using a subtitle parsing based approach. The short video clips for which the subtitle contained words related to emotions were recommended by the parser. The human labellers then chose or discarded the recommended clips and/or their emotion label. The AFEW data partitions are subject and movie independent, which results in different environment/scenes and more subjects. This experimentation protocol makes the task of emotion recognition in the wild non-trivial. The EmotiW 2013 event highlighted several challenges (e.g. robust facial part detection of non-frontal faces, temporal dynamics extraction, handling noise during speech analysis etc), which need to be tackled for affect analysis in the wild. Following EmotiW 2013, the second and third EmotiW challenges were organised to further address these challenges. EmotiW [1] added more data to the AFEW database, including more complex data to further stimulate research in affective computing towards real-world conditions. During EmotiW 2015 [5], a new sub-challenge—image based static facial expression recognition—was introduced, which was based on the Static Facial Expressions in the Wild (SFEW) database [3]. The SFEW database has been extracted from the AFEW database using a fiducial points based clustering technique. The papers published in this special issue were invited from teams from the three EmotiW challenges.

✉ Abhinav Dhall
abhinav.dhall@canberra.edu.au

Roland Goecke
roland.goecke@ieee.org

Tom Gedeon
tom.gedeon@anu.edu.au

Nicu Sebe
sebe@disi.unitn.it

[1] University of Canberra, Canberra, Australia

[2] Australian National University, Canberra, Australia

[3] University of Trento, Trento, Italy

1 Featured work

Kahou et al. [8] present a deep learning based approach for emotion recognition. This article is the extension of their winning entry [6] in the EmotiW 2013 challenge. The proposed technique consists of a convolutional neural network for face analysis, a bag of words for mouth area analysis and a deep belief network for audio signal analysis.

Liu et al. [13] discuss the extension of their winning entry [12] in the EmotiW 2014 challenge. A Riemannian kernel based approach is proposed and experiments with different

classifiers performed. In addition to the experiments on the AFEW database, the generalisation of the approach is also shown on the CK+ [14] database.

Sun et al. [17] propose a hierarchical classifier based approach for emotion recognition in the wild. An ensemble of standard visual features are extracted followed by feature and decision level fusions. Their approach [16] shows an increase in performance over the EmotiW 2014 baseline.

Kaya et al. [10] present a bimodal approach for emotion recognition based on extreme learning machine classification. Data augmentation is performed using various databases [9] and the performance of extreme learning machines is compared with that of partial least square based regression.

Kaechele et al. [7] extract audio, video and meta-features in their multimodal approach. Furthermore, the challenges in emotion recognition in real-world conditions are discussed. The results are presented for the EmotiW 2013 and 2014 datasets.

Zong et al. [18] propose a transfer learning based approach for facial expression recognition on the EmotiW 2015 image based sub-challenge data. The authors also conduct experiments on the audio only part of AFEW.

Kim et al. [11] propose a deep learning based approach for image based facial expression recognition. Exponential decision fusion is performed to infer the final expression label. The approach achieves the highest classification accuracy in the EmotiW 2015 image based facial expression recognition sub-challenge.

2 Challenges and discussion

The EmotiW challenge series aimed at providing a standard platform for emotion recognition researchers with a focus on the challenges presented by unconstrained environmental conditions as often found in real-world data, in comparison to the highly constrained environmental conditions in laboratory recorded data. This special issue and the EmotiW challenge series bring attention to the unaddressed but known obstacles in this field. The series has shown that emotion recognition in unconstrained conditions is difficult. At EmotiW 2015, the state-of-art classification accuracy performance has been 54 % for AFEW and 62 % for SFEW. This relatively low performance highlights the need for more research in this area.

The EmotiW challenges provide an opportunity for researchers to create a full end-to-end emotion recognition system. There are several problems at different stages in the emotion recognition pipeline, which the participants have tried to solve. One example in particular is that authors have tried to improve the facial parts detection performance, which is crucial for emotion recognition in the wild. Furthermore, authors have tried to address the problems of head pose movement, illumination, noise in audio and lack of labelled data. For analysing the effect of context in emotion recognition researchers have used the meta-data. Future challenges will carry forward the created platform and more data and newer problems (e.g. group-level emotion recognition) will be introduced.

Based on the papers in this special issue, it is clear that there is a need for large emotion recognition databases. As creating emotion related databases is a laborious task, effort is required in researching faster methods. The result discussions in several papers points out that classifying into seven emotion recognition classes can be ambiguous for some classes (e.g. disgust and fear). One possible direction is to use the compound expressions of emotions [15] based labelling. Another possibility is using multiple labels to define the emotion of a subject. Labelling large amounts of data is time consuming; however, with the advent in crowdsourcing, this task can be performed faster. The papers in the special issue focus on audio, video and meta-data modalities only. As part of future work, there is scope for use of other modalities such as EEG, eye gaze, skin galvanic response for emotion recognition in the wild. The main challenge here is how these modalities are captured in real-world environments.

The papers in the special issue present a glimpse of the state-of-art techniques in emotion recognition. The papers present an insight into the problem and we hope that the special issue will help in generating more interest into the problem of emotion recognition in the wild.

Acknowledgments We wish to thank the anonymous reviewers for their valuable comments and suggestions. We also wish to thank the Editor-in-chief Prof. Jean-Claude Martin and Springer staff Ms. Divyalochany Thangavel, who have helped in bringing together this special issue.

References

1. Dhall A, Goecke R, Joshi J, Sikka K, Gedeon T (2014) Emotion recognition in the wild challenge 2014: Baseline, data and protocol. In: Proceedings of the ACM on International conference on multimodal interaction (ICMI)
2. Dhall A, Goecke R, Joshi J, Wagner M, Gedeon T (2013) Emotion recognition in the wild challenge 2013. In: Proceedings of the ACM on international conference on multimodal interaction (ICMI), pp 509–516
3. Dhall A, Goecke R, Lucey S, Gedeon T (2011) Static facial expression analysis in tough conditions: data, evaluation protocol and benchmark. In: Proceedings of the IEEE international conference on computer vision and workshops BEFIT, pp 2106–2112
4. Dhall A, Goecke R, Lucey S, Gedeon T (2012) Collecting large, richly annotated facial-expression databases from movies. IEEE Multim 19(3):0034
5. Dhall A, Ramana Murthy O, Goecke R, Joshi J, Gedeon T (2015) Video and image based emotion recognition challenges in the wild: Emotiw 2015. In: Proceedings of the 2015 ACM on international conference on multimodal interaction, ACM, pp 423–426

6. Ebrahimi S, Pal C, Bouthillier X, Froumenty P, Jean S, Konda KR, Vincent P, Courville A, Bengio Y (2013) Combining modality specific deep neural networks for emotion recognition in video. In: Proceedings of the ACM on international conference on multimodal interaction (ICMI), pp 543–550
7. Kächele M, Schels M, Meudt S, Pam G, Schwenker F (2016) Revisiting the emotiw challenge: How wild is it really? J Multimodal User Interfaces. doi:10.1007/s12193-015-0202-7
8. Kahou SE, Bouthillier X, Lamblin P, Gulcehre C, Michalski, V, Konda K, Jean S, Froumenty P, Courville A, Vincent P et al (2016) Emonets: Multimodal deep learning approaches for emotion recognition in video. J Multimodal User Interfaces. doi:10.1007/s12193-015-0195-2
9. Kaya H, Salah AA (2014) Combining modality-specific extreme learning machines for emotion recognition in the wild. In: Proceedings of the 16th international conference on multimodal interaction, ACM, pp 487–493
10. Kaya H, Salah AA (2016) Combining modality-specific extreme learning machines for emotion recognition in the wild. J Multimodal User Interf 1–10
11. Kim BK, Lee H, Roh J, Lee SY (2016) Hierarchical committee of deep cnns with exponentially-weighted decision fusion for static facial expression recognition
12. Liu M, Wang R, Li S, Shan S, Huang Z, Chen X (2014) Combining multiple kernel methods on riemannian manifold for emotion recognition in the wild. In: Proceedings of the 16th international conference on multimodal interaction, ACM, pp 494–501
13. Liu M, Wang R, Shan S, Chen X (2016) Learning mid-level words on Riemannian manifold for action recognition. J Multimodal User Interfaces (to appear)
14. Lucey P, Cohn, JF, Kanade T, Saragih J, Ambadar Z, Matthews I (2010) The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression. In: Proceedings of the IEEE conference on computer vision and pattern recognition and workshops (CVPRW), pp 94–101
15. Martinez A, Du S (2012) A model of the perception of facial expressions of emotion by humans: Research overview and perspectives. J Mach Learn Res 13(1):1589–1608
16. Sun B, Li L, Zuo T, Chen Y, Zhou G, Wu X (2014) Combining multimodal features with hierarchical classifier fusion for emotion recognition in the wild. In: Proceedings of the 16th international conference on multimodal interaction, ACM, pp 481–486
17. Sun B, Li L, Zuo T, Chen Y, Zhou G, Wu X (2016) Combining feature-level and decision-level fusion in a hierarchical classifier for emotion recognition in the wild. J Multimodal User Interfaces. doi:10.1007/s12193-015-0203-6
18. Zong Y, Zheng W, Huang X, Yan K, Yan J, Zhang T (2016) Emotion recognition in the wild via sparse transductive transfer linear discriminant analysis. J Multimodal User Interf pp 1–10