

Emotion-aware cross-modal domain adaptation in video sequences

Citation for published version (APA):

Athanasiadis, C. (2022). *Emotion-aware cross-modal domain adaptation in video sequences*. [Doctoral Thesis, Maastricht University]. Maastricht University. <https://doi.org/10.26481/dis.20221020ca>

Document status and date:

Published: 01/01/2022

DOI:

[10.26481/dis.20221020ca](https://doi.org/10.26481/dis.20221020ca)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Summary: Emotion aware cross-modal domain adaptation in video sequences

Christos Athanasiadis

October 27, 2022

In his “Treatise of Human Nature”, David Hume argued that: “ Reason is, and ought only to be the slave of the passions!!”.

Emotion detection and recognition play a fundamental role in humans, especially in our everyday social engagement, when we are experiencing social behaviors (for example, in relationship maintenance). Moreover, they can prove really useful when addressing people in public (for instance, during a political debate or discussion). Not only do emotions play an important role in everyday life and social interactions of each person, but also they are crucial in human perception and cognition. Therefore, developing an automatic interface to rationalize about human emotions could prove to be very helpful, especially to those people that lack the ability to properly recognizing other people’s emotions.

At the same time, developing intelligent systems that are capable of recognizing emotions is an important, but a rather challenging task. Humans experience a wide range of emotions when interacting with an interface. Previous research has provided firm evidence that emotions vigorously affect motivation and hence play a crucial role when users interact with interfaces (such as learning or video games, for instance). However, it was not until recently that modern technology started to embody users’ affect analysis and not only use their conventional interaction methodology with systems (such as button clicks, feedback, etc.).

In this dissertation, a study of emotion recognition by employing sensorial modalities (such as auditory signals) with an insufficient amount of data is performed. Domain adaptation approaches are applied to leverage emotion-related information with great availability (such as facial expressions) and transfer this knowledge to emotion cues with less available information to increase their emotion recognition performance. In particular, we shed light on the audio-visual cross-modal relationship. Having a system performing emotion recognition from modalities such as audio (with limited access in audio features), the task is to leverage availability in face datasets (where there are more available data) with the purpose of enhancing the classification efficiency.

The main pillars of this research concern the implementation of metric learning and generative models to facilitate the task of cross-modal domain adaptation between audio and face within the affect context. Furthermore, an extensive study in the ways that temporal dynamics between audio and face modalities could contribute to domain adaptation is employed. Finally, we facilitate also the inverse task that is to perform face-based emotion recognition by eliciting auditory source information.