

Accepted Manuscript

Modeling and Synthesis of Kinship Patterns of Facial Expressions

Itır Önal Ertuğrul, László A. Jeni, Hamdi Dibeklioglu



PII: S0262-8856(18)30157-4
DOI: doi:[10.1016/j.imavis.2018.09.012](https://doi.org/10.1016/j.imavis.2018.09.012)
Reference: IMAVIS 3721
To appear in: *Image and Vision Computing*
Received date: 16 October 2017
Revised date: 12 May 2018
Accepted date: 12 September 2018

Please cite this article as: Itır Önal Ertuğrul, László A. Jeni, Hamdi Dibeklioglu , Modeling and Synthesis of Kinship Patterns of Facial Expressions. *Imavis* (2018), doi:[10.1016/j.imavis.2018.09.012](https://doi.org/10.1016/j.imavis.2018.09.012)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Modeling and Synthesis of Kinship Patterns of Facial Expressions

Itr Önal Ertuğrul^{a,*}, László A. Jeni^a, Hamdi Dibeklioglu^b

^aRobotics Institute, Carnegie Mellon University, Pittsburgh, PA, USA

^bDepartment of Computer Engineering, Bilkent University, Ankara, Turkey

Abstract

Analysis of kinship from facial images or videos is an important problem. Prior machine learning and computer vision studies approach kinship analysis as a verification or recognition task. In this paper, for the first time in the literature, we propose a kinship synthesis framework, which generates smile and disgust videos of (probable) children from the expression videos (smile and disgust) of parents. While the appearance of a child's expression is learned using a convolutional encoder-decoder network, another neural network models the dynamics of the corresponding expression. The expression video of the estimated child is synthesized by the combined use of appearance and dynamics models. In order to validate our results, we perform kinship verification experiments using videos of real parents and estimated children generated by our framework. The results show that generated videos of children achieve higher correct verification rates than those of real children. Our results also indicate that the use of generated videos together with the real ones in the training of kinship verification models, increases the accuracy, suggesting that such videos can be used as a synthetic dataset. Furthermore, we evaluate the expression similarity between input and output frames, and show that the proposed method can fairly retain the expression of input faces while transforming the facial identity.

Keywords: Kinship synthesis, Kinship verification, Temporal analysis, Facial Action Units, Facial dynamics

1. Introduction

Analysis of kin relations from facial appearance has gained popularity in recent years. This research topic has several potential applications including missing child/parent search, social media analysis, family album organization, and image annotation [1]. Majority of prior studies in kinship analysis focus on *kinship verification* [2, 3, 4]; given a pair of face images, they try to identify whether these two have a kin relationship or not. Contrarily, *kinship recognition* studies aim to classify the type of kin relationship such as Father-Daughter, Mother-Son, etc. [5].

In addition to general appearance of face, style and appearance of expressions can also be inherited. Facial expressions of congenitally blind and deaf children with phocomelia, who are incapable of sensing their relatives' face by touching, are shown to be similar to those of their parents [6]. Moreover, [7] reports that a blind-born son, who was abandoned by his mother two days after birth, displays similar facial expressions with the biological mother. Findings of [4] show that the use of expression dynamics extracted from videos together with facial appearance leads to more accurate kinship verification compared to employing only facial appearance. Thus, although facial expressions may comprise learned characteristics, it is clear that they are at least partially inherited.

All of the previous studies approach the kinship analysis as a verification or recognition problem. They model the underlying relationship between a pair of images or videos, yet, what

these models learn is not visible to the naked eye. In this study, for the first time in the literature, we focus on *kinship synthesis*, and generate facial expression videos of children using the expression video of their parents. Kinship synthesis has several benefits. First of all, since we synthesize videos, the hereditary patterns inherited from parent to child can be observed by humans. Observed patterns may even be useful for genetic research. Secondly, there are only two kinship video/expression databases (UvA-NEMO Smile [8] and UvA-NEMO Disgust databases [9]) available for automatic kinship analysis, thus, our models can be used to create synthetic kinship videos for further research. Lastly, with the help of our model, people will be able to preview how their (probable) future child may look like, as well as seeing his/her smile/disgust expression as a video. Therefore, if a child, whose appearance and expression dynamics are unknown, has been missing for years, generated videos of him/her (based on expressions of the parents) would be better references for the search compared to pictures drawn by forensic artists.

This study is the very first exploration of synthesizing facial images and expression videos for a kin relationship. By transforming temporal dynamics and appearance of a given subject, we generate a video of his/her probable children. Furthermore, we show that the synthesized samples can be used to improve the state of the art in kinship verification.

We extend our previous study [10] in many ways. Along with an extended literature, (1) we use intensity of facial action units (AUs) instead of facial landmark displacement for both expression matching and learning temporal dynamics, (2) we model the facial appearance in a holistic manner, rather than

*Corresponding author

Email address: iertugru@andrew.cmu.edu (Itr Önal Ertuğrul)

Download English Version:

<https://daneshyari.com/en/article/11030081>

Download Persian Version:

<https://daneshyari.com/article/11030081>

[Daneshyari.com](https://daneshyari.com)