Dataset Bias Exposed in Face Verification

**Titulo**  
Dataset Bias Exposed in Face Verification

**Autores**  
Eric López-López, Xosé M. Pardo, Carlos V. Regueiro, Roberto Iglesias, Fernando E. Casado

**Tipo**  
Artículo de revista

**Fonte**  

**Rank**  
Ranked Q1 in Software by SJR

**ISSN**  
2047-4938

**DOI**  
10.1049/iet-bmt.2018.5224

**Abstract**  
Most facial verification methods assume that training and testing sets contain independent and identically distributed samples, although, in many real applications, this assumption does not hold. Whenever gathering a representative dataset in the target domain is unfeasible, it is necessary to choose one of the already available (source domain) datasets. In this paper, a study was performed over the differences among six public datasets, and how this impacts on the performance of the learned methods. In the considered scenario of mobile devices, the individual of interest is enrolled using a few facial images taken in the operational domain, while training impostors are drawn from one of the public available datasets. This work tried to shed light on the inherent differences among the datasets, and potential harms that should be considered when they are combined for training and testing. Results indicate that a drop in performance occurs whenever training and testing are done on different datasets compared to the case of using the same dataset in both phases. However, the decay strongly depends on the kind of features. Besides, the representation of samples in the feature space reveals insights into to what extent bias is an endogenous or an exogenous factor.