Attribute-based quality assessment for demographic estimation in face videos

Fabiola Becerra-Riera, Annette Morales-González and Heydi Méndez-Vázquez
Advanced Technologies Application Center (CENATAV), Havana, Cuba
{fbecerra,amorales,hmendez}@cenatav.co.cu

Jean-Luc Dugelay
Digital Security Department, EURECOM, France
{Jean-Luc.Dugelay@eurecom.fr}

1 Introduction

Demographic soft biometrics (e.g. gender, age, ethnicity) are among the most frequently used traits for improving and complementing the performance of biometric systems [1].

Most existing works regarding facial demographic estimation are focused on still image datasets, although nowadays the need to analyze video content in real applications is increasing.

We propose a pipeline for the automatic estimation of gender, ethnicity and age in videos.

Our main contribution is to use an attribute-specific quality assessment procedure to select most relevant frames from a video sequence for each of the three demographic modalities. Selected frames are classified with fine-tuned MobileNet models [2] and a final video prediction is obtained with a majority voting strategy.

2 Proposal

Quality Assessment

We associate the relevance of a frame within a sequence to 12 quality parameters relatives to Pose, Illumination, Occlusion, Resolution, Sharpness, Mouth State, Eyes State, Gaze, Color Leveling, Face Centering, Red Eyes and Uniform Background.

Some quality measures could have more or less impact over the relevance of a frame, depending on the specific demographic attribute to classify.

We employed Random Forest (RF) classifiers [3] to learn the relations between the quality measures and each classification task (gender, age, ethnicity).

The output score of the 12 quality measures were concatenated and the resulting features for good and bad classification samples were used to train each RF quality classifier.

Final video prediction was obtained with a majority voting strategy among best quality frames selected by the RF classifier.

Demographic Estimators

We used fine-tuned MobileNet models to make the real-time demographic estimation of the selected video frames as efficient as possible.

Training was performed on three publicly uncontrolled image datasets: IMDB-Wiki Dataset, UTKFace Dataset and LFW.

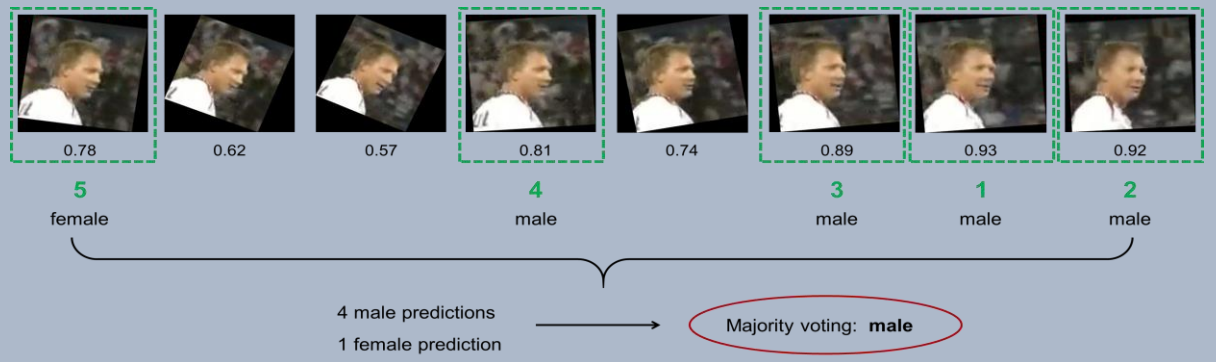
Formalization

Given a sequence of video frames $F = \{f_1, f_2, \dots f_m\}$ and an attribute α with $L = \{l_1, l_2, \dots l_k\}$ possible labels, we define $q^n \subseteq F$ as the set of the n best quality frames of the sequence $(1 \leq n \leq m)$ and $q^n_{l_k} \subseteq q^n$ as the set of the $f \in q^n$ for which the classification according to α (from now on defined as C_α) corresponds to the l_k label:

$$\boldsymbol{q^n_{l_k} = \{\mathbf{f} \in q^n \mid C_\alpha(\mathbf{f}) = l_k} \qquad (1)$$

Then, the l_i resulting after applying the majority voting strategy over the sequence F given the α attribute, responds to the following formulation:

$$\boldsymbol{l_i \in L\,(1 \leq i \leq k) \mid \forall l_j \in L\,(1 \leq j \leq k), j \neq i), \left|q^n_{l_j}\right| < \left|q^n_{l_i}\right|} \qquad (2)$$



Attribute-based quality assessment for demographic estimation in face videos

Fabiola Becerra-Riera, Annette Morales-González and Heydi Méndez-Vázquez
Advanced Technologies Application Center (CENATAV), Havana, Cuba
{fbecerra,amorales,hmendez}@cenatav.co.cu

Jean-Luc Dugelay
Digital Security Department, EURECOM, France
{Jean-Luc.Dugelay@eurecom.fr}

3 Experiments

Dataset Selection

1. **UvA-Nemo:**
Fairly good quality video collection with gender and age annotations, created to analyze the change in smile dynamics across different ages.

2. **EURECOM Augmented:**
Dataset representing frames of a video, fully annotated with demographic data and augmented with added noise to simulate a video scenario with several frames of different qualities.

3. **Youtube Faces (YTF):**
Uncontrolled video collection for which we mapped the Labelled Faces in the Wild (LFW) gender and ethnicity labels to its corresponding identities.

➢ Gender, ethnicity and age distribution:

Datasets	Gender		Ethnicity				Age			
	F	M	C	Af	As	I/L	(0-18)	(19-30)	(31-59)	(60-)
UvA-Nemo	185	215	400	-	-	-	150	81	150	19
EURECOM	14	38	20	3	11	18	-	45	7	-
YTF	149	285	315	33	15	72	-	-	-	-

We performed experiments in the selected datasets by comparing several frame combination strategies:

- Individual frames: Considers frames as single independent images.
- Sequence (all frames): Performs a majority voting among all frames in a sequence.
- Sequence - quality N frames: Performs the majority voting on the N top relevant frames.
- Sequence - random N frames: Performs the majority voting on N random frames.

Attribute-based quality assessment for demographic estimation in face videos

Fabiola Becerra-Riera, Annette Morales-González and Heydi Méndez-Vázquez
Advanced Technologies Application Center (CENATAV), Havana, Cuba
{fbecerra,amorales,hmendez}@cenatav.co.cu

Jean-Luc Dugelay
Digital Security Department, EURECOM, France
{Jean-Luc.Dugelay@eurecom.fr}

Results and Discussion

Gender

➢ Gender classification results in the “Deliberate” and “Spontaneous” subsets from UvA-Nemo dataset, and also in the “Entire” collection:

Classifier	Strategy	Deliberate Accuracy (%)				Spontaneous Accuracy (%)				Entire dataset Accuracy (%)			
		Overall	G-Mean	< 20	> 19	Overall	G-Mean	< 20	> 19	Overall	G-Mean	≤ 20	> 20
MobileNet (Ours)	Individual frames	90.27	89.11	83.90	94.64	89.25	**88.62**	**83.61**	93.92	89.59	88.47	83.20	94.08
	Sequence (all frames)	90.51	88.84	83.29	94.77	88.96	87.89	82.21	93.97	**89.76**	88.54	83.32	**94.09**
	Sequence - random 5 frames	**91.28**	**89.69**	**84.35**	**95.37**	88.46	87.45	81.94	93.32	89.19	87.83	82.12	93.94
	Sequence - random 10 frames	90.67	88.95	83.29	95.00	**89.29**	88.12	82.44	94.20	89.59	88.31	82.89	**94.09**
DEX [30]	Individual frames	88.75	87.68	83.12	92.49	87.96	86.73	79.72	94.35	88.19	86.75	80.61	93.36
	Sequence (all frames)	89.72	88.50	83.71	93.56	88.09	86.52	79.08	94.66	88.95	87.51	81.68	93.75
	Sequence - random 5 frames	89.72	87.83	82.50	93.50	87.59	86.06	78.47	94.38	88.87	87.39	81.43	93.79
	Sequence - random 10 frames	89.73	87.56	80.80	94.89	88.09	86.49	78.76	**94.98**	88.79	87.24	81.05	93.91
Dantcheva and Brémond [17]	Sequence (all frames)	-	84.53	76.92	92.89	-	84.58	76.92	93	-	-	-	-
Bilinski *et al.* [32]	Sequence (all frames)	-	-	-	-	-	-	-	-	-	**88.62**	**86.30**	91.01

- We were not able to train a gender quality estimator for this collection due to the lack of bad quality samples and their slight differences with the good quality ones.
- The experiments allows to compare our baseline method to other state-of-the-art algorithms.

➢ Gender classification results in EURECOM and YTF datasets:

Classifier	Strategy	EURECOM Accuracy (%)				YTF Accuracy (%)			
		Overall	G-Mean	Female	Male	Overall	G-Mean	Female	Male
MobileNet (Ours)	Individual frames	76.18	81.07	94.84	69.30	94.30	92.24	86.97	97.83
	Sequence (all frames)	91.35	92.88	96.43	89.47	94.66	92.67	86.63	99.15
	Sequence - random 5 frames	84.62	88.86	**100.0**	78.95	94.55	92.75	87.23	98.64
	Sequence - random 10 frames	87.50	91.04	**100.0**	82.89	94.55	92.67	86.93	98.81
	Sequence - quality 5 frames	**100.0**	**100.0**	**100.0**	**100.0**	95.75	94.29	**89.67**	99.15
	Sequence - quality 10 frames	99.04	99.34	**100.0**	98.68	**95.86**	**94.37**	**89.67**	**99.32**
DEX [30]	Individual frames	89.35	77.81	60.58	99.95	91.57	89.01	82.66	95.84
	Sequence (all frames)	92.31	84.52	71.43	**100.0**	90.73	88.13	80.55	96.43
	Sequence - random 5 frames	91.35	82.38	67.86	**100.0**	90.84	88.16	81.46	95.41
	Sequence - random 10 frames	92.31	84.52	71.43	**100.0**	91.38	89.09	82.32	96.43
	Sequence - quality 5 frames	**99.04**	**98.20**	**96.43**	**100.0**	92.49	90.27	84.37	**96.60**
	Sequence - quality 10 frames	98.08	96.36	92.86	**100.0**	**92.87**	**91.03**	**85.91**	96.46

- The proposed quality assessment is effective without dependence on the dataset or the classifier.
- 100% of classification accuracy in EURECOM.
- The quality assessment strategy favored the results of the minority class (Females).

Attribute-based quality assessment for demographic estimation in face videos

Fabiola Becerra-Riera, Annette Morales-González and Heydi Méndez-Vázquez
Advanced Technologies Application Center (CENATAV), Havana, Cuba
{fbecerra,amorales,hmendez}@cenatav.co.cu

Jean-Luc Dugelay
Digital Security Department, EURECOM, France
{Jean-Luc.Dugelay@eurecom.fr}

Results and Discussion

Ethnicity

- Ethnicity classification results in EURECOM and YTF datasets:

		EURECOM Accuracy (%)						YTF Accuracy (%)					
Classifier	Strategy	Overall	G-Mean	Caucasian	African	Asian	Other	Overall	G-Mean	Caucasian	African	Asian	Other
MobileNet (Ours)	Individual frames	63.57	68.84	57.69	91.98	66.67	63.48	75.67	68.55	83.20	89.91	82.84	35.64
	Sequence (all frames)	85.58	89.25	72.50	**100.0**	95.45	91.67	76.63	67.89	83.90	87.67	76.67	37.67
	Sequence - random 5 frames	75.00	79.22	75.00	**100.0**	72.73	72.22	76.30	65.49	84.20	87.67	70.00	35.62
	Sequence - random 10 frames	79.81	84.17	77.50	**100.0**	86.36	75.00	76.20	65.11	84.20	89.04	70.00	34.25
	Sequence - quality 5 frames	**99.04**	**99.37**	**97.50**	**100.0**	**100.0**	**100.0**	**78.59**	**70.19**	**85.54**	**93.15**	**76.67**	**39.73**
	Sequence - quality 10 frames	97.12	98.07	92.50	**100.0**	**100.0**	**100.0**	78.37	69.85	85.39	**93.15**	**76.67**	39.04

- We were not able to find an available state-of-the-art pre-trained model for this task.
- By using the quality strategy, in the EURECOM dataset:
  - The minority classes Asian and Other achieved 100% of accuracy, showing great improvements.
  - The majority class Caucasian also was largely favored.

Attribute-based quality assessment for demographic estimation in face videos

Fabiola Becerra-Riera, Annette Morales-González and Heydi Méndez-Vázquez
Advanced Technologies Application Center (CENATAV), Havana, Cuba
{fbecerra,amorales,hmendez}@cenatav.co.cu

Jean-Luc Dugelay
Digital Security Department, EURECOM, France
{Jean-Luc.Dugelay@eurecom.fr}

Results and Discussion

Age

- MAE in the estimation of exact age in UvA-Nemo dataset:

Classifier	Strategy	MAE (years) Overall	G-Mean	0-9	10-19	20-29	30-39	40-49	50-59	60-69	70-79
MobileNet (Ours)	Frames	4.94 (± 0.76)	3.92	0.62	2.49	7.23	7.02	6.98	6.75	6.6	2.27
	Seq. (all frames)	4.88 (± 0.73)	3.77	**0.53**	2.55	7.27	**6.6**	6.59	7.2	6.32	2.08
	Seq. random 5	4.95 (± 0.81)	3.79	0.55	2.52	7.25	6.81	6.81	6.83	6.43	2.09
	Seq. random 10	4.93 (± 0.75)	3.76	**0.53**	2.54	7.17	6.69	6.81	7.12	6.23	2.05
DEX [30]	Frames	4.13 (± 0.88)	3.19	1.33	1.43	6.21	6.98	5.89	3.99	4.76	1.15
	Seq. (all frames)	**4.09 (± 1.03)**	**3.02**	1.23	1.37	6.18	6.74	5.69	3.91	**4.69**	**0.95**
	Seq. random 5	4.18 (± 0.98)	3.23	1.22	1.41	6.13	7.23	5.82	3.94	4.85	1.38
	Seq. random 10	4.09 (± 1.06)	3.11	1.18	**1.35**	6.31	6.74	5.74	**3.78**	4.87	1.22
Dibeklioğlu *et al.* [16]	Seq. (all frames)	4.81 (± 4.87)	5.96	2.73	2.99	**5.45**	6.83	**4.35**	8.45	10.87	13.18
Number of samples	-	1 240	-	158	333	215	171	250	66	30	17

- We were not able to train an age quality estimator for this collection, as explained before.
- DEX classifier was slightly more accurate for age groups over 50 years old; however, DEX is no suitable for real-time video applications due to its larger processing time.

- MAE in the estimation of exact age in EURECOM dataset:

Classifier	Strategy	MAE (years) Overall	G-Mean	25	26	27	28	29	30	31	32	33	36	38
MobileNet (Ours)	Frames	7.19	8.36	6.31	6.90	6.22	6.42	7.60	8.05	12.85	9.04	7.05	13.20	12.20
	Seq. (all frames)	5.56	6.86	4.22	5.86	3.75	4.64	9.83	5.70	10.50	8.17	5.50	14.00	10.00
	Seq. random 5	6.13	6.91	3.57	6.14	4.71	6.77	6.33	7.30	14.00	9.00	**5.00**	**10.50**	**8.00**
	Seq. random 10	5.45	6.74	2.50	5.07	4.21	4.45	7.17	8.10	13.50	8.33	5.50	14.50	10.50
	Seq. quality 5	4.36	**5.56**	2.50	4.07	3.33	3.95	**5.83**	**4.10**	**9.00**	**6.67**	5.50	13.50	11.00
	Seq. quality 10	**4.21**	5.57	**2.43**	**3.50**	**3.25**	**3.18**	6.17	4.50	11.00	7.00	5.50	14.00	11.00
DEX [30]	Frames	9.58	10.01	9.24	8.53	9.00	9.82	9.54	10.50	12.57	11.10	10.74	11.26	8.59
	Seq. (all frames)	8.87	8.23	6.64	9.07	7.42	9.73	7.33	10.2	12.00	12.67	14.00	17.00	**1.00**
	Seq. random 5	7.93	8.73	7.86	**6.64**	6.54	8.59	7.50	8.70	12.00	10.00	8.50	10.00	11.50
	Seq. random 10	8.91	9.49	**6.36**	7.86	7.21	10.45	10.50	10.20	12.00	12.00	14.00	12.50	5.50
	Seq. quality 5	**5.78**	**4.11**	7.14	7.57	5.29	**6.59**	8.33	**2.80**	**6.50**	**3.33**	**2.00**	3.00	**1.00**
	Seq. quality 10	6.59	4.57	7.21	7.50	**4.38**	9.68	**5.83**	6.60	7.00	5.83	**2.00**	**2.50**	**1.00**

- DEX classifier showed better performance in the classification of people over 30 years old; our MobileNet was more accurate in the classification of the younger.

Attribute-based quality assessment for demographic estimation in face videos

Fabiola Becerra-Riera, Annette Morales-González and Heydi Méndez-Vázquez
Advanced Technologies Application Center (CENATAV), Havana, Cuba
{fbecerra,amorales,hmendez}@cenatav.co.cu

Jean-Luc Dugelay
Digital Security Department, EURECOM, France
{Jean-Luc.Dugelay@eurecom.fr}

4 Conclusions

The quality strategy works with different classifiers and under different conditions, allowing:

- ✓ Less number of frames to be classified.
- ✓ Less processing time.
- ✓ Improved estimation accuracy.
- ✓ Bias mitigation on specific gender, ethnicity and age.

5 Future Work

- Explore other video frame combination beyond majority voting.
- Deeper analysis regarding quality problems affecting specific demographic attributes.

References

[1] P. Tome, J. Fierrez, R. Vera-Rodriguez, and M. S. Nixon, “Soft biometrics and their application in person recognition at a distance,” Trans. Info. For. Sec., vol. 9, no. 3, pp. 464–475
[2] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, “Mobilenets: Efficient convolutional neural networks for mobile vision applications,” CoRR, vol. abs/1704.04861, 2017.
[3] L. Breiman, “Random forests,” Mach. Learn., vol. 45, no. 1, pp. 5–32, Oct. 2001.