

# BoT-FaceSORT: Bag-of-Tricks for Robust Multi-Face Tracking in Unconstrained Videos (Supplementary Material)

Jonghyeon Kim<sup>1\*</sup>, Chan-Yang Ju<sup>1\*</sup>, Gun-Woo Kim<sup>2</sup>, and Dong-Ho Lee<sup>1†</sup>

<sup>1</sup> Department of Applied Artificial Intelligence, Hanyang University, South Korea  
{jonghyeon, karunogi, dhlee72}@hanyang.ac.kr

<sup>2</sup> Department of Computer Science and Engineering, Gyeongsang National University, South Korea  
gunwoo.kim@gnu.ac.kr

<https://github.com/bellhyeon/BoT-FaceSORT>

## A MovieShot Dataset

In this section, we provide the details of our proposed dataset for multi-face tracking, MovieShot. It consists of 10 movie clip videos from YouTube.

As shown in Table 1, MovieShot contains between 1,284 and 5,086 frames per clip, 9 to 58 unique tracks, 12 to 86 shot changes, and 1,979 to 14,247 face detections. The dataset consists of movies produced in the following five countries: India (IN), the United Kingdom (UK), the United States (US), South Korea (KR), and Hong Kong (HK). As illustrated in Figure 1, our dataset features faces from diverse ethnicities, including American, European, Indian, Korean, and Chinese. This diversity ensures that the dataset can be used to develop and evaluate unbiased tracking algorithms.

Table 1: Basic Statistics of the MovieShot dataset

Video	Frames	Tracks	Shot changes	Face detections	Country
# 1	4405	34	65	12997	IN
# 2	1284	10	26	1979	UK
# 3	3156	46	35	3426	US
# 4	2172	34	12	4847	US
# 5	2986	41	86	3549	KR
# 6	4340	58	85	9821	UK
# 7	3862	13	35	5877	KR
# 8	2558	37	36	4603	HK
# 9	5086	9	74	7858	US
# 10	2703	9	13	14247	US

\* Major in Bio Artificial Intelligence

† Corresponding Author



Fig. 1: Sample videos of the MovieShot dataset: (a) Shaolin Soccer, HK; (b) The Truman Show, US; (c) Charlie and the Chocolate Factory, UK; (d) The Scam, KR; (e) 3 Idiots, IN; (f) Les Misérables, UK.