

# Mise-en-Scène Dataset: Stylistic Visual Features of Movie Trailers

## SUMMARY

---

This dataset provides a set of **7 low-level VISUAL features** extracted from **13373 movie trailers**. The movie IDs are in agreement with the movie IDs provided by "MovieLens dataset", the FULL version as in July 18, 2016, containing in total 22M ratings for 33K movies. All the movie titles, ratings and associated movie genres and tags can be collected from the MovieLens website.

## INFORMATION ABOUT THE DATASET

---

This dataset provides a set of 7 low-level VISUAL features extracted from 13373 movie trailers. The data are contained in **three files**:

- LLM\_sual\_Features13K\_Log.csv
- LLM\_sual\_Features13K\_Quantile.csv
- LLM\_sual\_Features13K\_QuantileLog.csv

More details about the contents and use of these files are as follows. The description of each column and each low-level visual feature is provided in the table below:

**Table 1: Description of columns**

Column #	Column Name	Description
1	ML_ID	MovieLens movie ID
2	f1	Average shot length
3	f2	Mean of color variance across the key Frames
4	f3	Standard deviation of color variance across the key Frames
5	f4	Mean of motion average across all the frames
6	f5	Mean of motion standard deviation across all the frames
7	f6	Mean of lighting key across the key frames
8	f7	Number of shots

## DESCRIPTION OF FILES:

---

1. LLM\_sual\_Features13K\_Log.csv

This file provides the low-level visual features presented in Table 1 scaled by passing them through a "Natural Logarithm" function. This changed the distributions to be approximately normal, as the original features in the dataset had a distribution similar to log-normal distribution.

## 2. LLM\_sual\_Features13K\_Quantile.csv

This file provides the low-level visual features presented in Table 1 normalized by applying "Quantile Normalization". This would change the distribution of all the features to be similar.

## 3. LLM\_sual\_Features13K\_QuantileLog.csv

This file provides the low-level visual features presented in Table 1 normalized by passing them through a "Natural Logarithm" function and then a "Quantile Normalization" scheme.

After applying the above normalization schemes, the features were scaled to the range [0-1].

## CITATION

=====

To acknowledge the use of the dataset in publications, please cite one of the following papers:

*Yashar Deldjoo, Mehdi Elahi, Paolo Cremonesi "How to Combine Visual Features with Tags to Improve the Movie Recommendation Accuracy", Springer – Proceedings of the 17h International conference on Electronic Commerce and Web Technologies (Ec-Web 2016), Porto, Portugal, September 5-8, 2016*

*Yashar Deldjoo, Mehdi Elahi, Paolo Cremonesi "Using Visual Features and Latent Factors for Movie Recommendation", ACM RecSys Workshop on New Trends in Content-based Recommender Systems (CBRecSys), ACM RecSys 2016, Massachusetts Institute of Technology (MIT), September 15-19, 2016*

## FURTHER INFORMATION ABOUT POLIMI@RECSYS

=====

**POLIMI@RECSYS** is a research group at the Department of Electronics, Information and Bioengineering (DEIB) at Politecnico di Milano in Milan, Italy. The research focus of POLIMI@RECSYS is mainly on Recommender Systems, in particular Multimedia Recommender Systems.

The dataset can be downloaded at: [https://www.researchgate.net/publication/305682269\\_Mise-en-Scene\\_Dataset\\_Stylistic\\_Visual\\_Features\\_of\\_Movie\\_Trailers\\_dataset](https://www.researchgate.net/publication/305682269_Mise-en-Scene_Dataset_Stylistic_Visual_Features_of_Movie_Trailers_dataset)

Please visit our homepage for further information: <http://recsys.deib.polimi.it>

Yashar Deldjoo: <http://www.ydeldjoo.me>

Mehdi Elahi: <http://www.linkedin.com/in/mehdielahi>

Paolo Cremonesi: <http://www.deib.polimi.it/eng/people/details/159156>