



EXPLORING “ARTIST IMAGE” USING CONTENT-BASED ANALYSIS OF PROMOTIONAL PHOTOS

Jānis Lībeks
jlibeks1@swarthmore.edu

Douglas Turnbull
turnbull@cs.swarthmore.edu

Motivation:

Long before Michael Jackson made music videos for MTV, and even before Elvis played The Ed Sullivan Show, the outward appearance, or *image*, of artists has played an important role in shaping how their music is received by audiences.

We explore how image similarity is related to music similarity by analyzing promotional photos of artists.

Our goal is to annotate artists with meaningful tags using a state-of-the-art image annotation system [1].

Dataset:

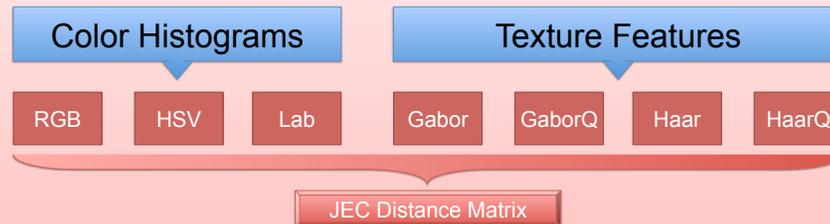
Obtained from



1. 50 most popular music genre tags
50 tags
2. 50 most popular artists for each tag
1704 artists
3. 5 most popular images for each artist
8483 images

Approach:

Use image similarity measure proposed by Makadia et al. [1].



Joint Equal Contribution (JEC): for each pair of images, take normalized average of seven distances.

1. For each image of the query artist, find the 1-nearest neighbor image (e.g., most similar “looking” artist)
2. Propagate tags to the query artist from set of similar artists by averaging their annotation vectors

Example: Nickelback

| Artist | rock | indie | classical | pop | punk |
|----------------------------------|------|-------|-----------|-----|------|
| Disturbed | 1 | 0 | 0 | 0 | 0 |
| Plain White T's | 1 | 1 | 0 | 1 | 1 |
| 311 | 1 | 1 | 0 | 0 | 1 |
| Wilco | 1 | 1 | 0 | 0 | 0 |
| Predicted Tags from JEC | 1 | .75 | 0 | .25 | .5 |
| Truth tags for Nickelback | 1 | 1 | 0 | 0 | 0 |

Evaluation:

Best performance with 4 images per artist.
37 out of 50 tags perform better than random.

| Tag | AUC* |
|---------------------|------|
| random | 0.50 |
| metal | 0.70 |
| death metal | 0.69 |
| melodic death metal | 0.68 |
| thrash metal | 0.67 |
| hardcore | 0.65 |
| trance | 0.64 |
| dance | 0.64 |
| heavy metal | 0.63 |
| indie pop | 0.63 |

*AUC – Area under ROC curve

Conclusions:

Images **can** be used to predict *some* genres.

Higher-level object detection might be used to improve performance (e.g., “10-gallon hat” or “cowboy boot” detectors for *country* music.)

References:

[1] A. Makadia, F. Pavlovic, S. Kumar. A new baseline for image annotation. ECCV, 2008.