

Data

Training data:

1000 images from 6 themes.

- architecture
- beach
- nature
- desert
- wildlife
- winter landscape

Test data:

Up to 100 images of 44 countries.

Features:

- $7500(\text{RGB}) + 288(\text{HOG}) = 7788$ components
- PCA: 18 components

Data is gathered from Instagram through the API[1]

Why 18 components in PCA?

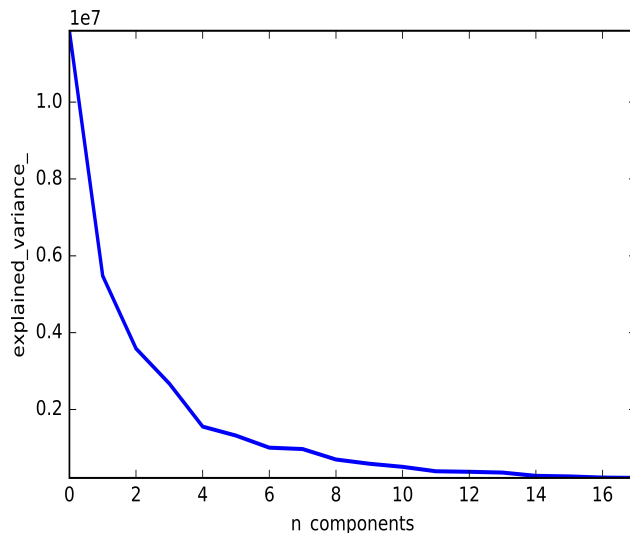
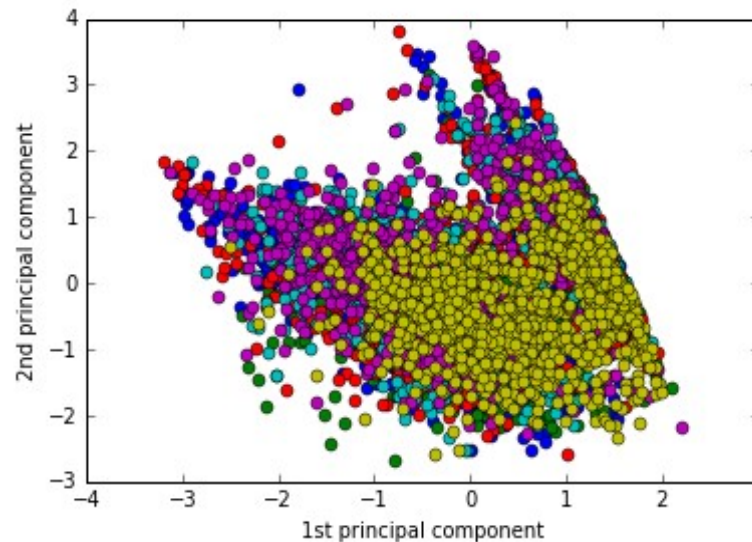


Figure showing number of components vs variance. The variance is almost zero beyond 16 components.

Why non-linear algorithms?



Plot of first two principal components suggesting non-linearity of data.

[1] <https://www.instagram.com/developer/>

Pre-processing

Training data



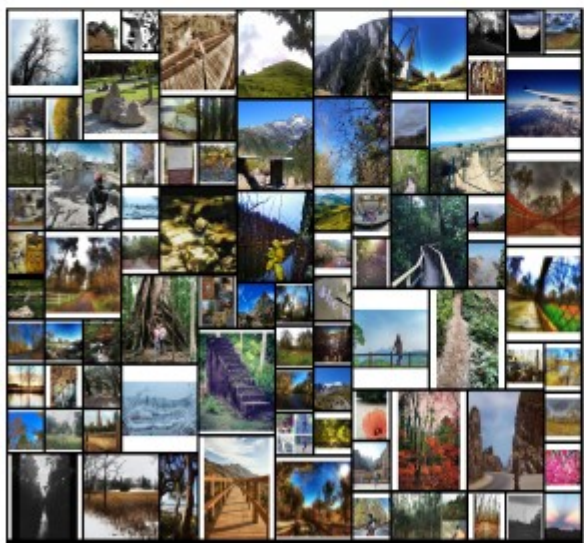
Architecture



Beach



Desert



Nature



Wildlife



WinterLandscape

Test data



Australia



Canada



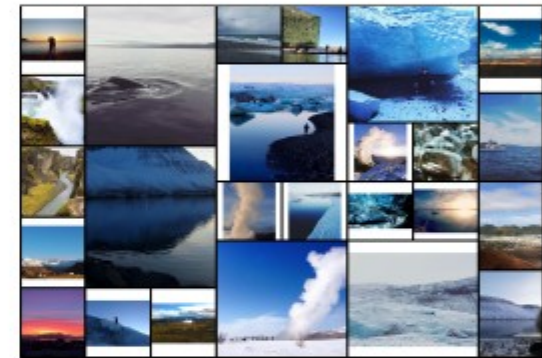
Egypt



Germany



Greece



Iceland



India

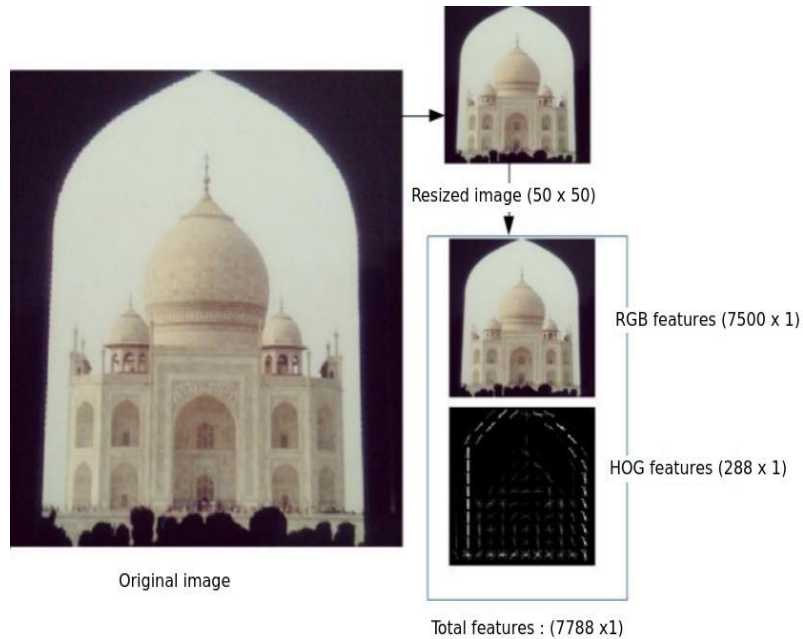


New Zealand



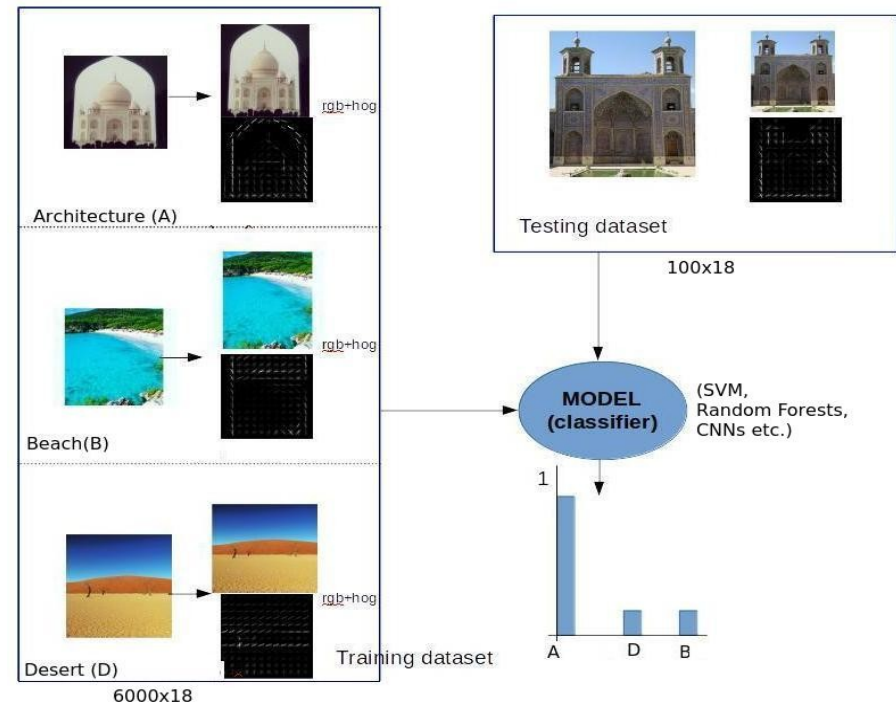
Tanzania

Feature extraction (explained in figure)



Feature extraction process explained in detail. Each image is first resized to 50x50 and HOG features are extracted. The combined feature is $(7500(\text{RGB}) + 288(\text{HOG})) = 7788$. To reduce dimensionality PCA is used and the size is reduced to 18.

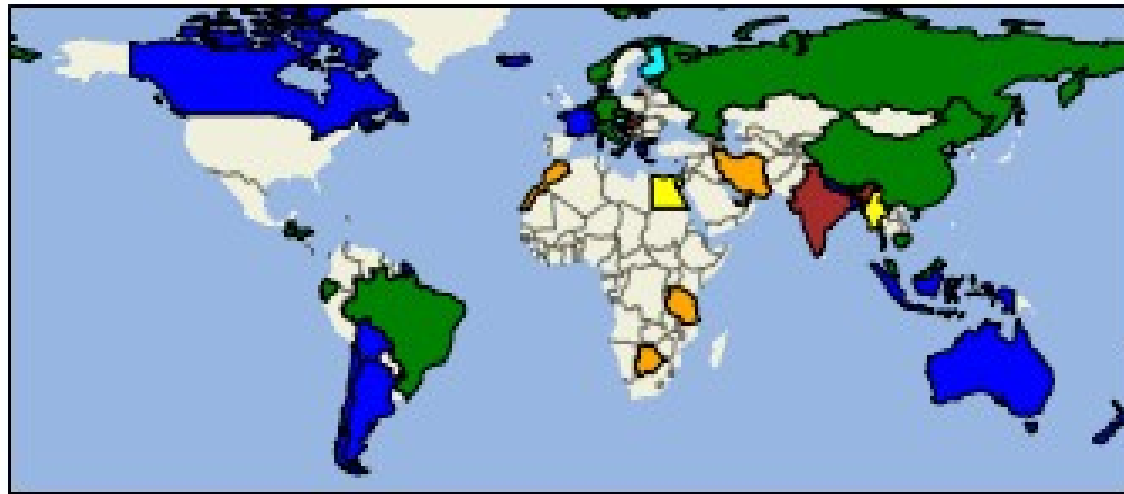
Proposed framework



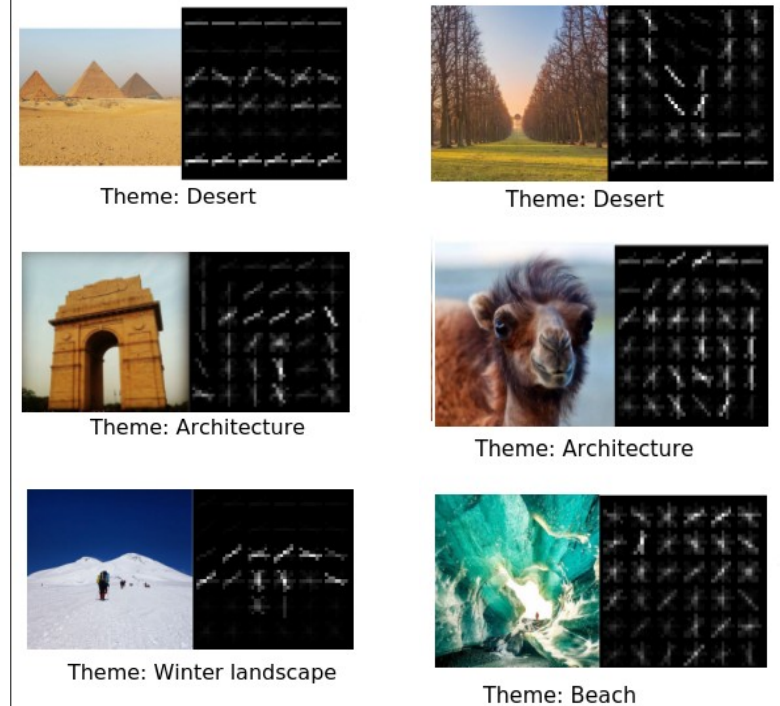
Training dataset of themes is trained using various classification methods. Test images are obtained from 44 countries. Output is probability distribution of tourist themes.

Tested method	Parameters	Accuracy
SVM	Kernel: RBF, C=2	53.41 %
Random Forests	Number of trees = 100	50 %

Table showing accuracy of the machine learning methods used



Maximum likelihood tourist theme for various destinations.



Example of correct classification

Analysis and Results