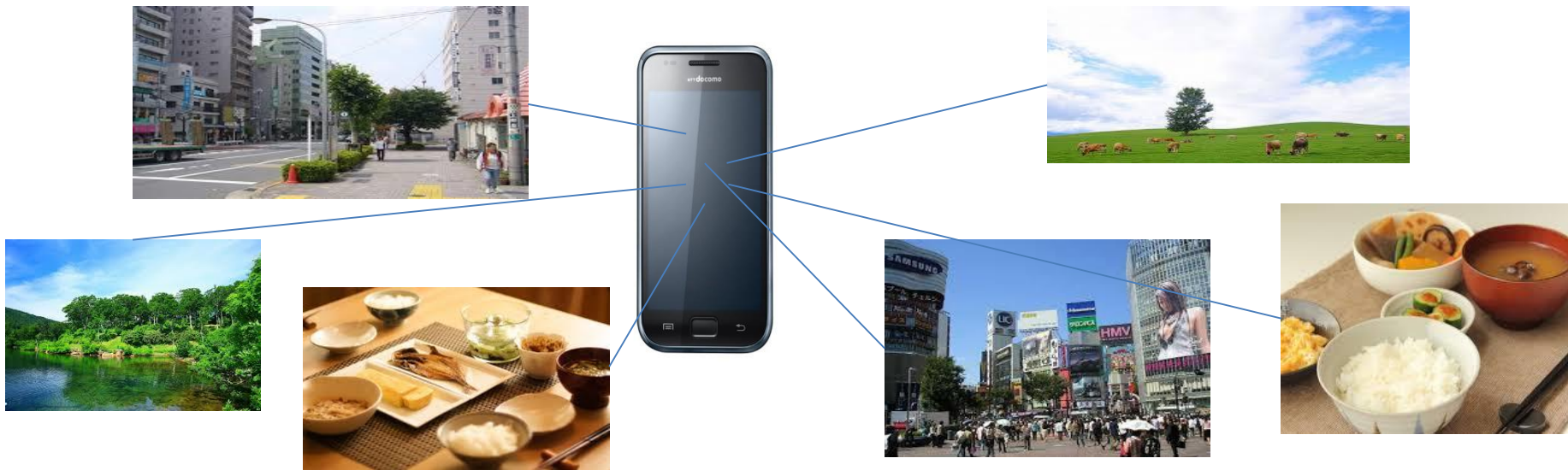


A Large-scale Analysis of Regional Tendency of Twitter Photos Using Only Image Features

Tetsuya Nagano, Takumi Ege, Wataru Shimoda and Keiji Yanai
Department of Informatics, The University of Electro-Communications, Tokyo

Introduction

- Many people to post photos on open SNSs such as Twitter and Instagram

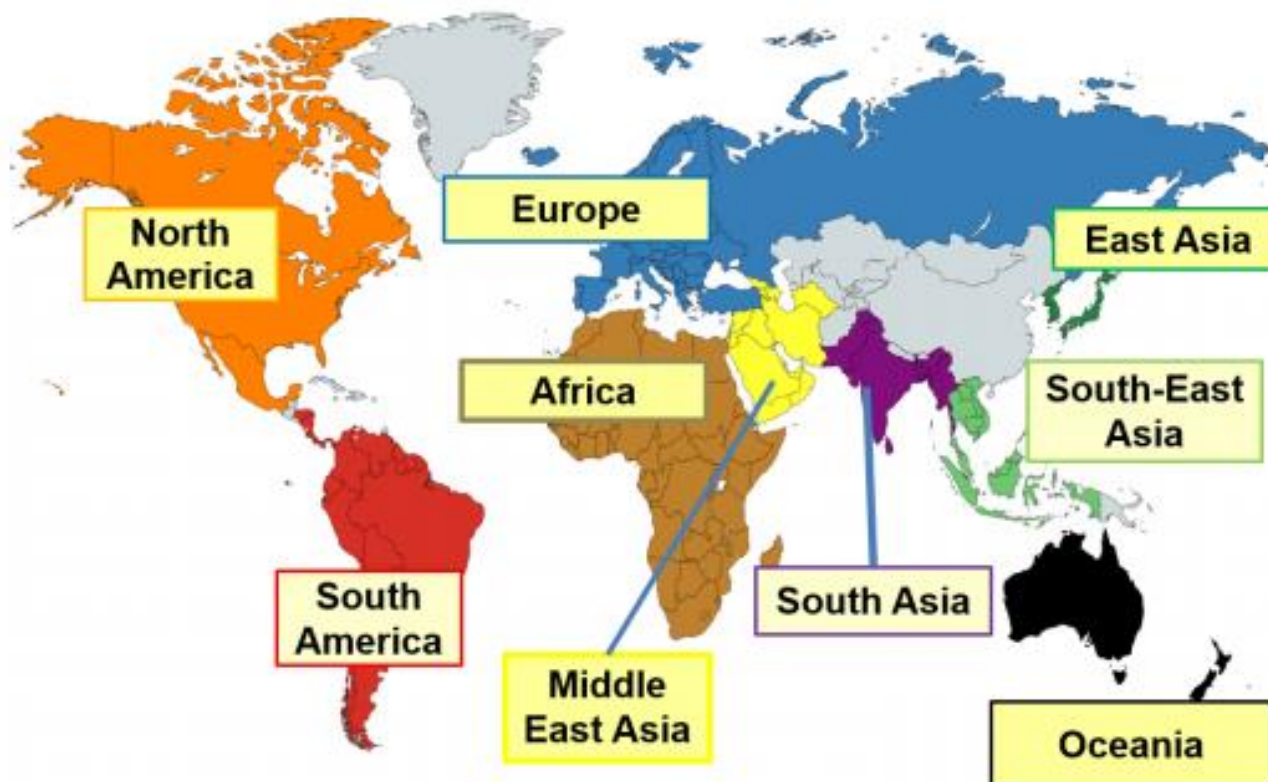


Introduction

- Flickr
 - Textual information attached to photos for photo sharing
- Twitter, Instagram
 - Do not tend to represent the contents of photos directly
 - Not for search but for explaining additional information

Introduction

- Analyze the differences on idea of the privacy issue at SNSs depends on culture and history



Related Works

- Typical works on Twitter photo,
 - Event detection
 - Using both text analysis and image analysis
- Our work
 - We analyze a million-scale of Twitter photos without using any textual information
 - We aim to detect the differences of regional tendency of posted photos to Twitter

Overview of the Proposed Method

- (1) Gather a million-scale of geotagged photos from the Twitter stream
- (2) Extract image features
- (3) Carry out clustering of them
- (4) Classify only large clusters
- (5) Compare the ratios of five categories between eight regions over the world

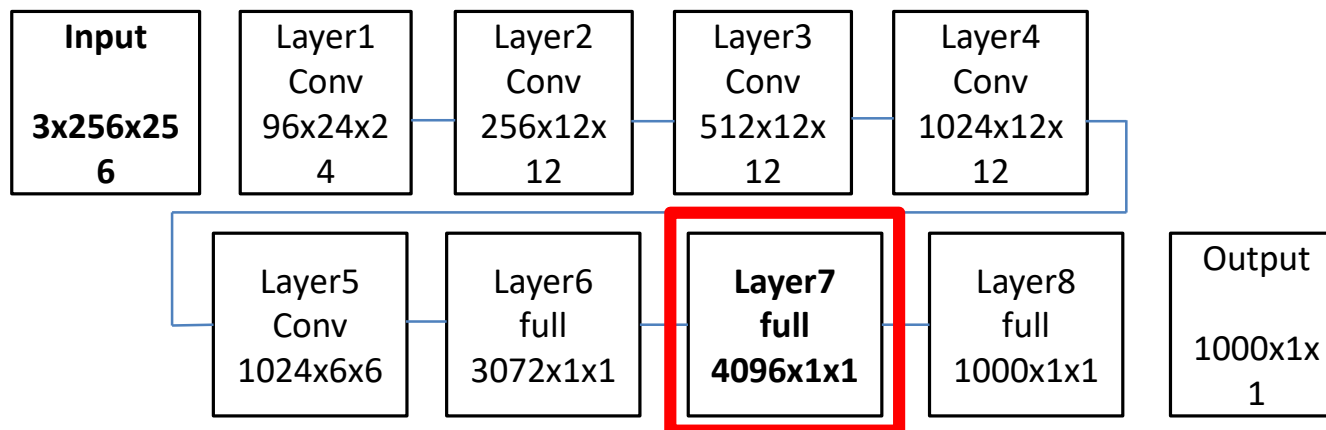
Detail of the Method

- We are continuously gathering geotagged photo tweets from the Twitter stream
- Two million geotagged Twitter photos which we had gathered for half years in 2016.



Image Classification

- AlexNet (CaffeNet)
 - Extract CNN features of 1000 images per one minutes
- FC6 layer of CaffeNet is 4096
- Principal Component Analysis (PCA)
 - 4096-d features into 128-d features



Clustering

- K-means clustering.
 - $k = 100$
- A small pre-liminary experiments with 1000 Twitter photos
 - To examine the difference on clustering results between the cases with and without PCA-based feature compression

Clusteling results (4096-d)



Clusteling results (128-d)



©2016/12/20/16



©2016/12/20/16



©2016/12/20/16

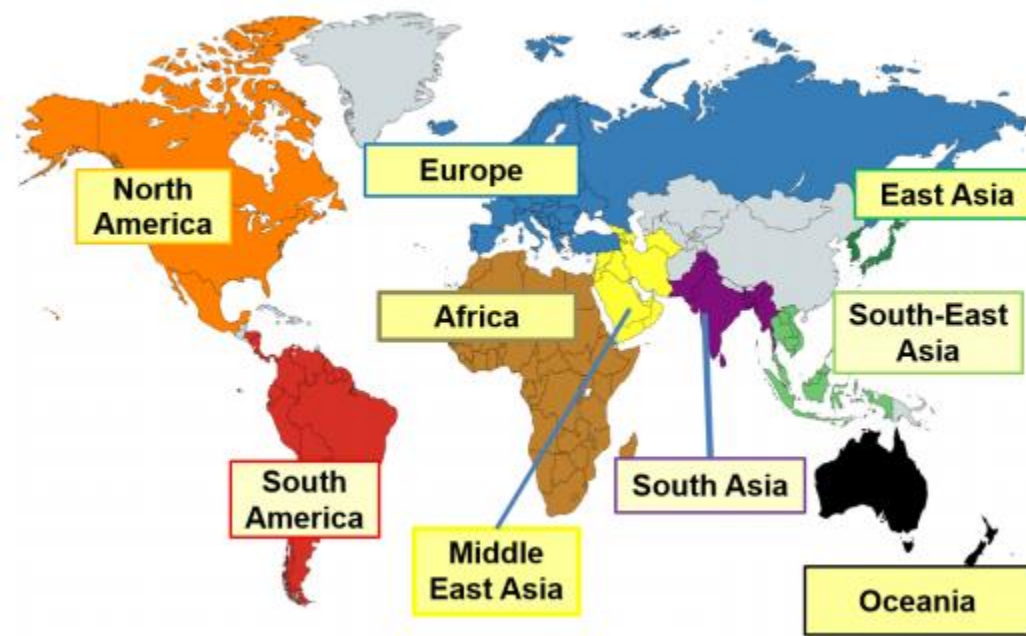


5 Experiments

- Dataset
 - Collected from January to June in 2016
 - 2,161,000 geotagged Twitter images
- Feature
 - CNN(128-d compressed by PCA)
- K-means
 - K-means with one-tenth images
 - Assigned rest of images into the nearest clusters
 - K=100

5.3 How to Analyze

- East Asia, North America, South America, Europe, Africa, Middle East, South Asia and South-East Asia, Oceania



5.3 How to Analyze

- pre-selected photo genres.
 - “people”
 - “building”
 - “document”
 - “scene”
 - “food”



Figure 5. “Food” in East Asia.



Figure 6. “Building” in North America.

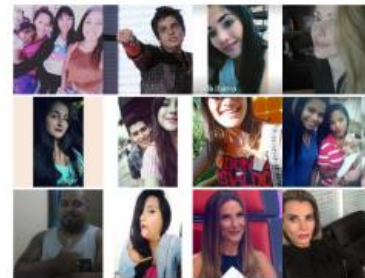
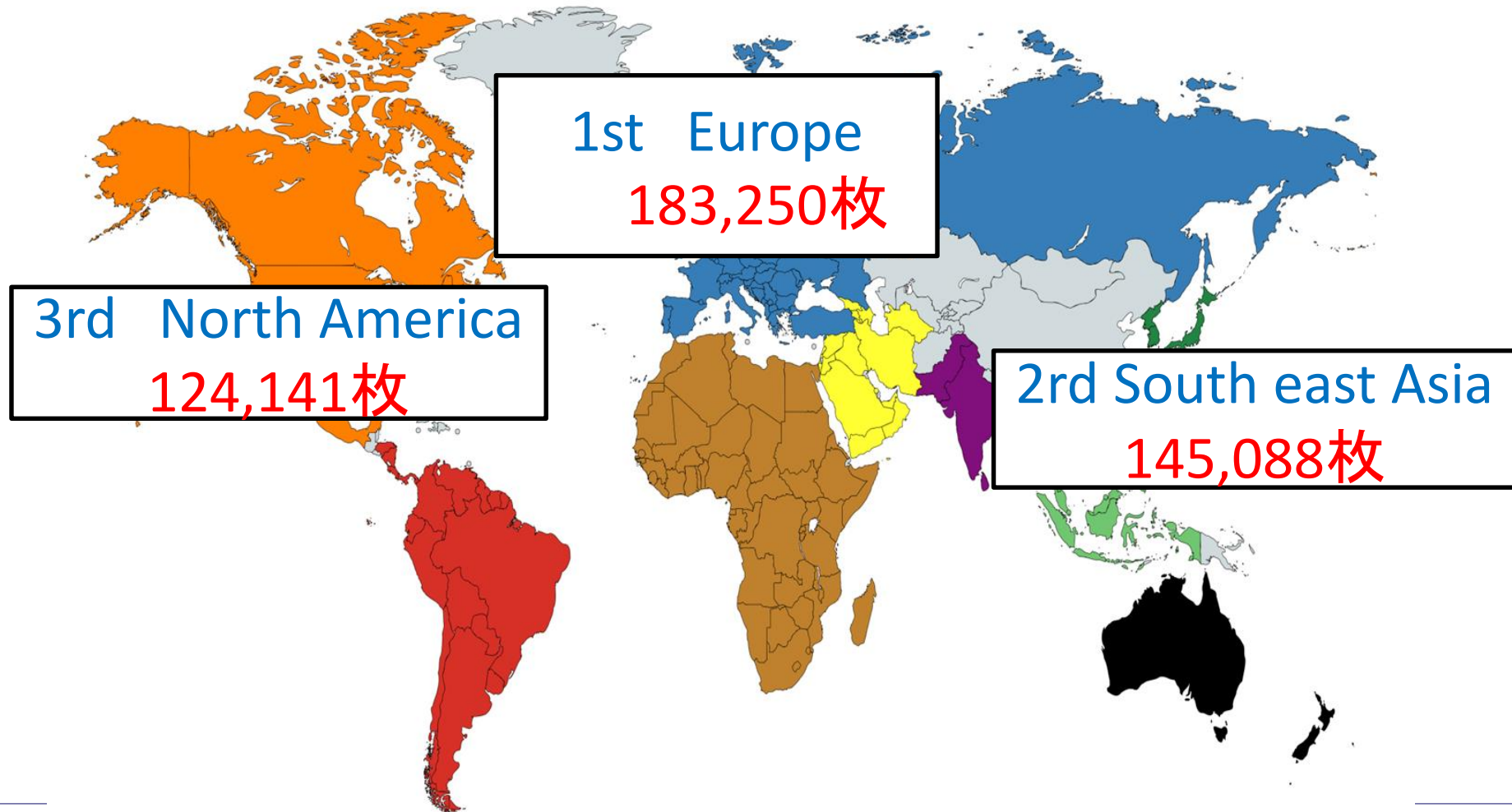


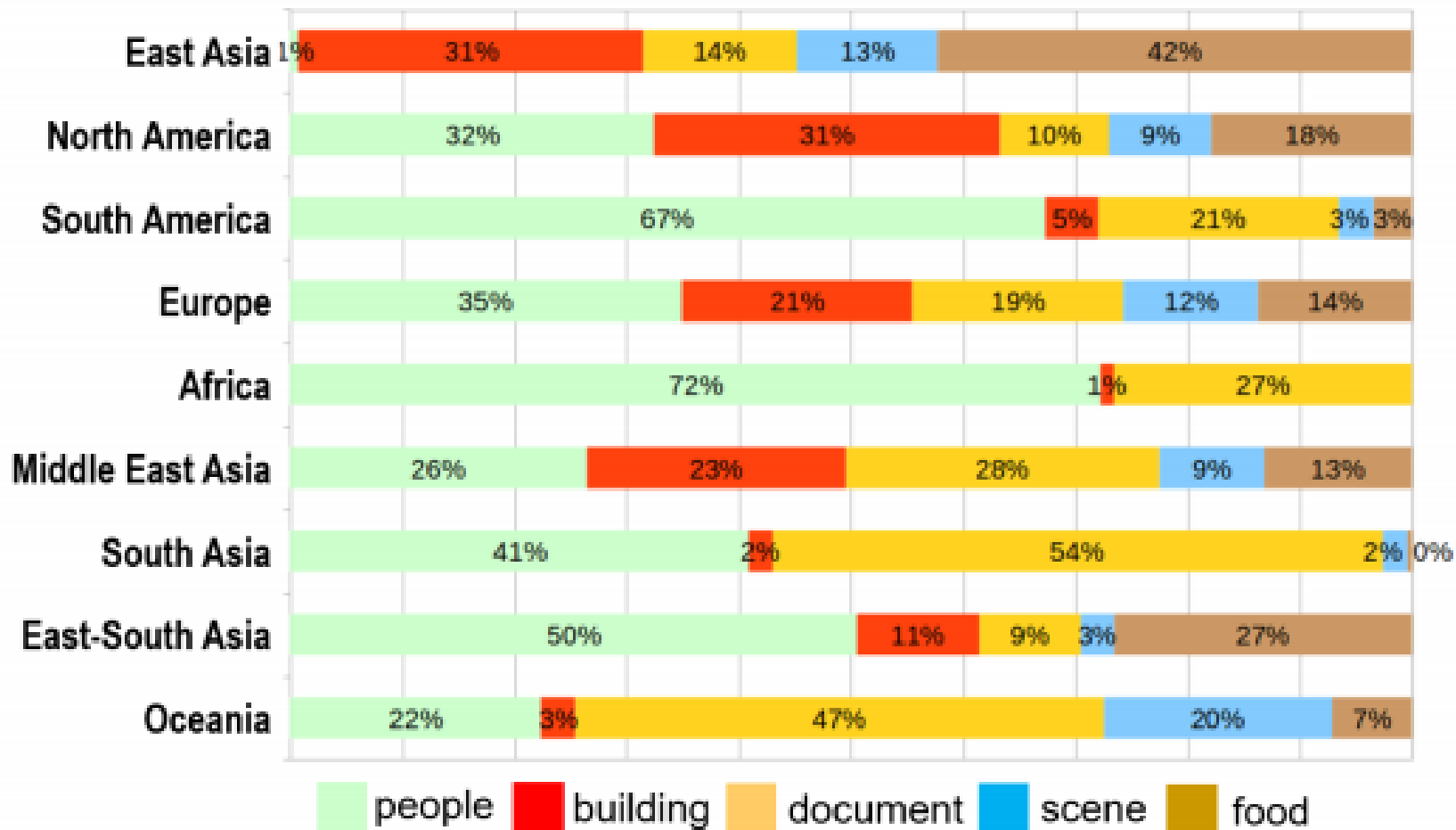
Figure 7. “People” in South America.



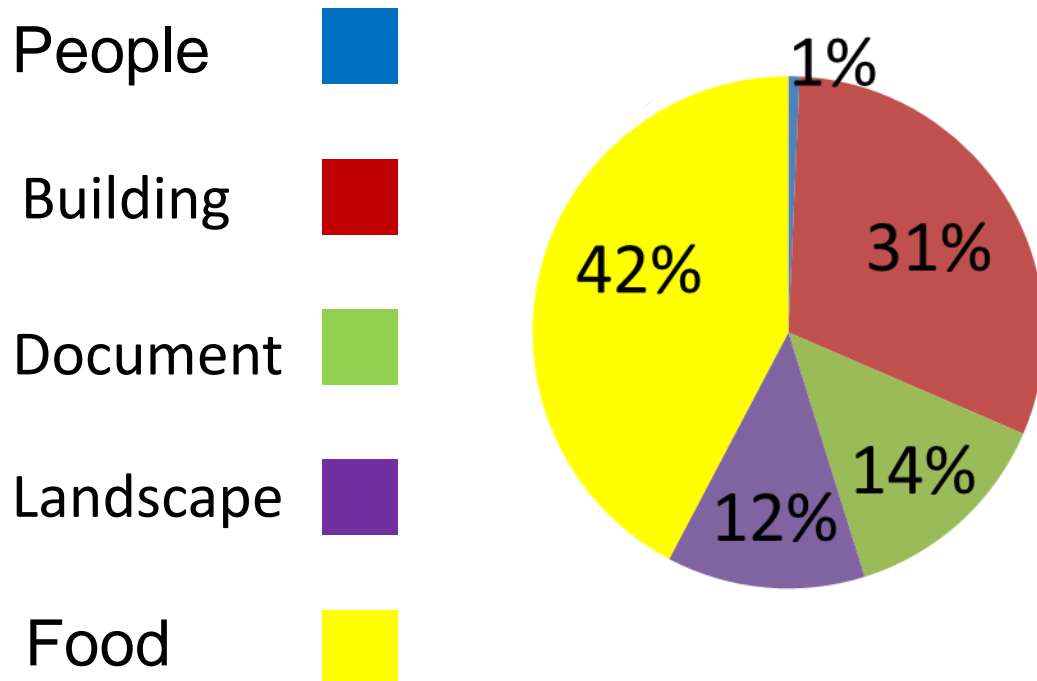
Figure 8. “Document” in Middle East.

Analysis of Regional Tendency of Photo Genres



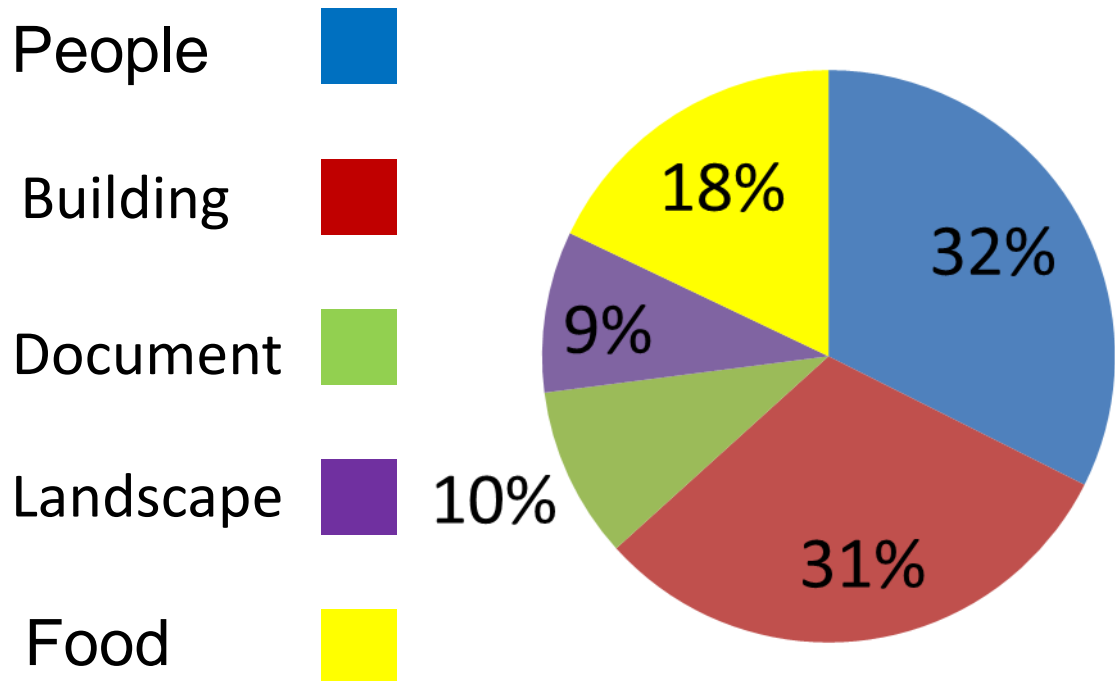


- East Asia
 - No people photos
 - Many building and food photos
 - The total ratio of building and food photos were more than 70%



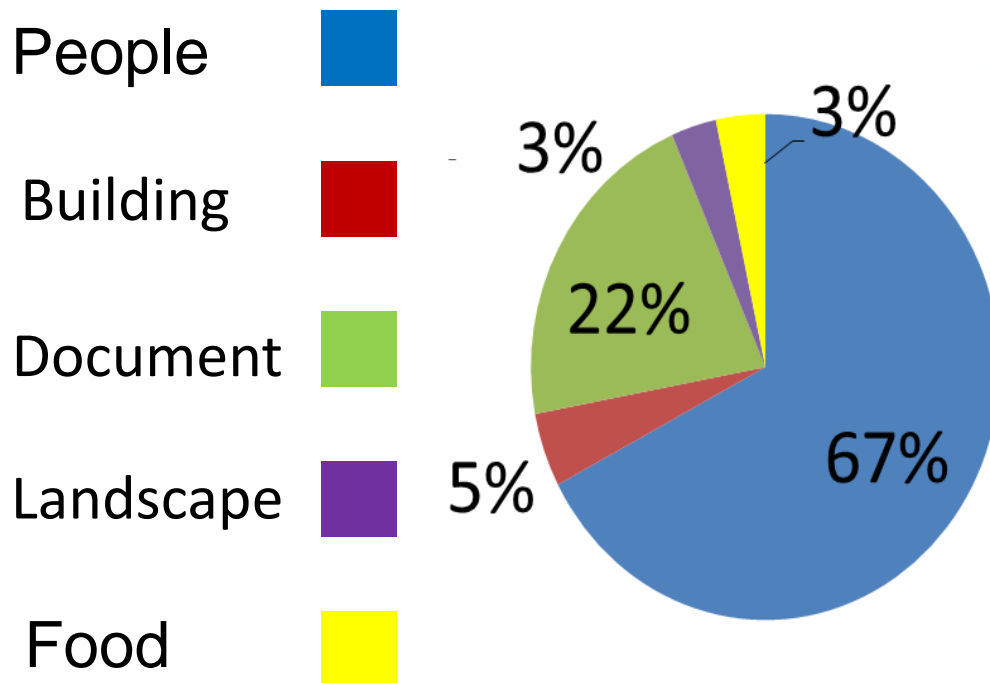
- North America

- The ratio of people and building were high more than 60%.



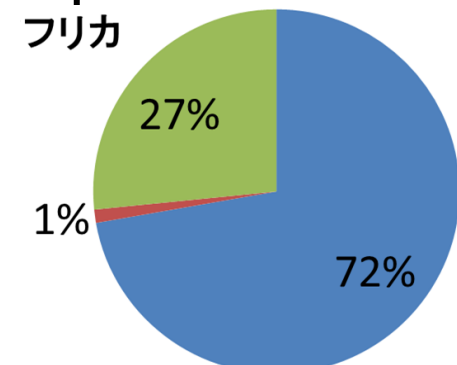
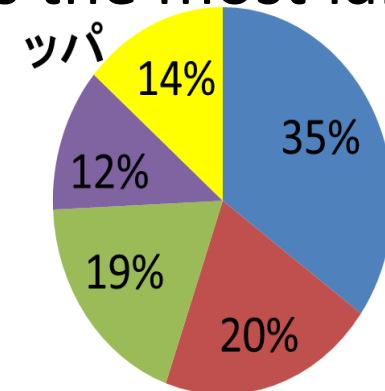
- South America

- People photos are the most popular genre (67%)



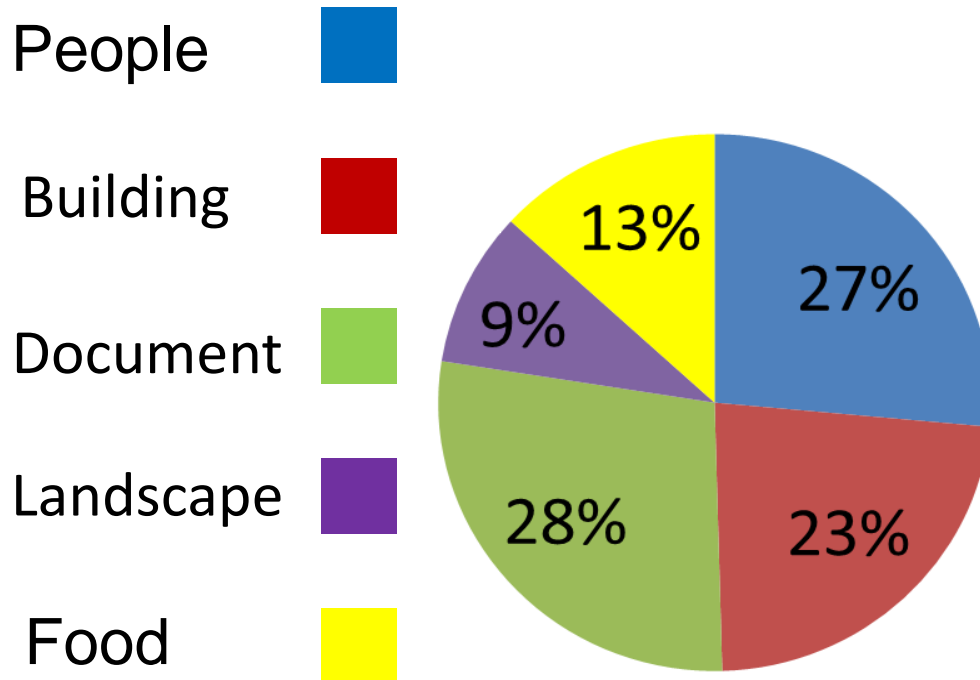
Analysis of Regional Tendency of Photo Genres

- Europe
 - The number of posted photos was the most large
 - The genres were well balanced.
- Africa
 - Almost no building, scene and food photos were posted
 - People photos occupied 70%.



- Middle East

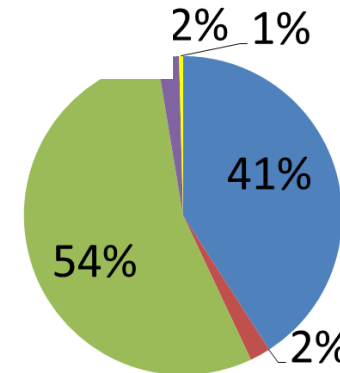
- Although the number of posts were fewer than Europe, all the five genres were balanced as well



Analysis of Regional Tendency of Photo Genres

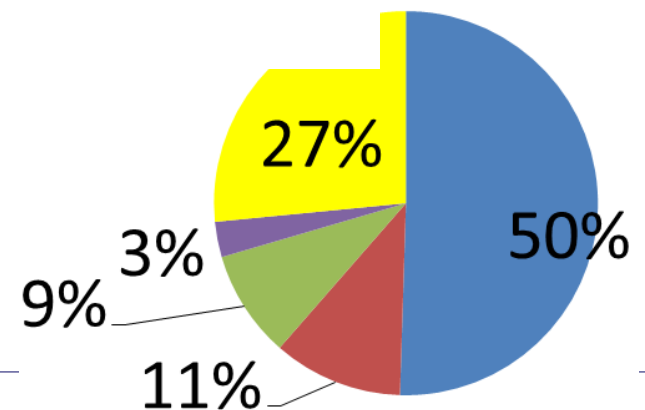
- South Asia

- More than half of the photos were document photos.
- This tendency was not observed in other regions

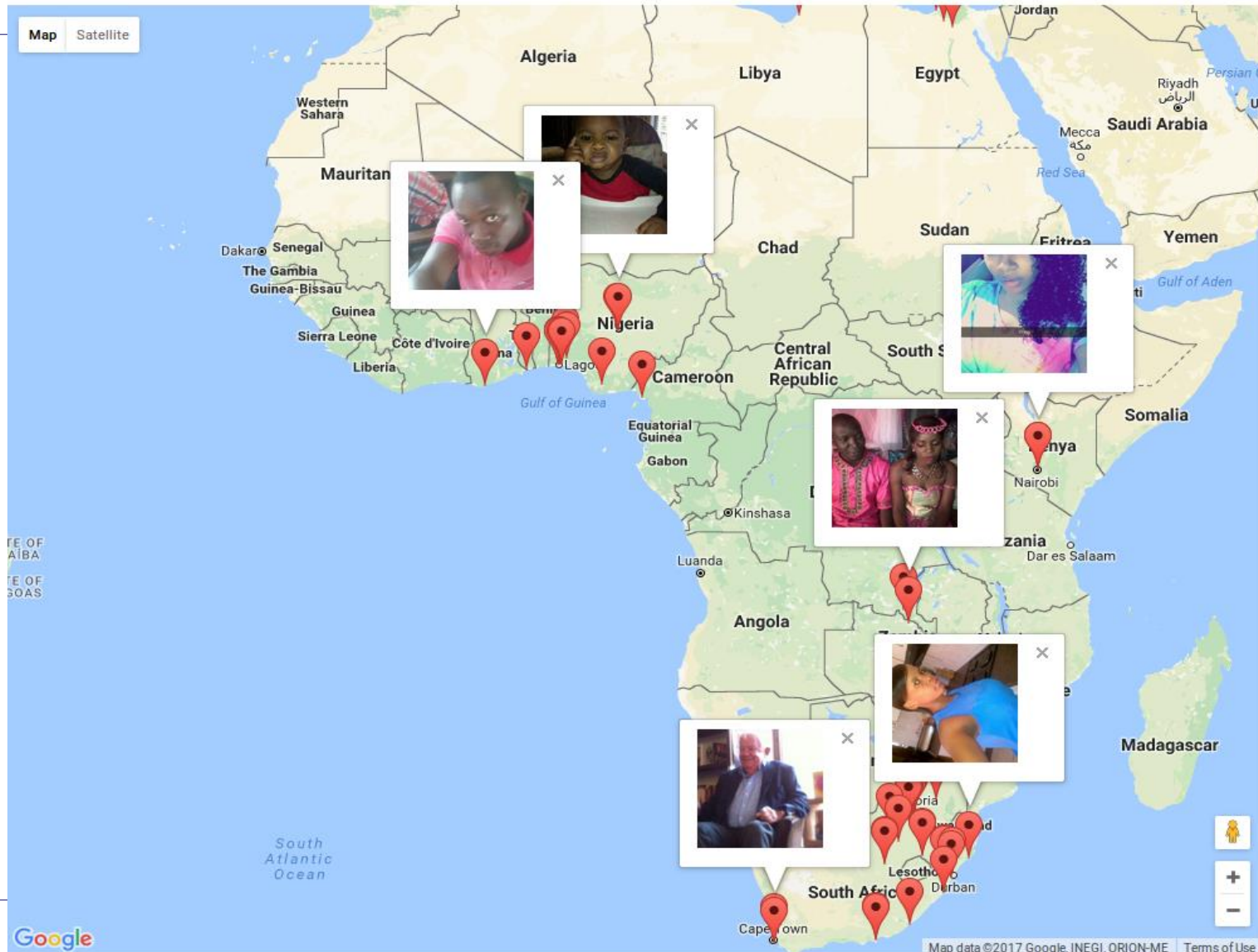


- SouthEast Asia

- People photos are the most and in addition food photos was the second most



Africa : People



Discussion

- Tendency
 - East Asia and East-South Asia,
 - Food photos are relatively high
 - South America, South Asia and East-South Asia
 - people photos are exceptionally high.
 - Europe and MiddleEast
 - well balanced.
- East Asia enjoy posting food photos
- South America, South Asia and EastSouth Asia like to post people photos without caring privacy issue.

Conclusions

- Analyzed the differences of the tendency of the photo genres of posted photos.
 - Gathered geotagged Twitter photos
 - Extracted CNN features
 - Carried out clustering
- Future works
 - making typical genres more fine-grained
 - “people photo” -> “selfy” and “group photo”

