Spatio-Temporal Sentiment Hotspot Detection Using Geotagged Photos

Yi Zhu and Shawn Newsam
Electrical Engineering & Computer Science
University of California at Merced
yzhu25,snewsam@ucmerced.edu

ABSTRACT

We perform spatio-temporal analysis of public sentiment using geotagged photo collections. We develop a deep learning-based classifier that predicts the emotion conveyed by an image. This allows us to associate sentiment with place. We perform spatial hotspot detection and show that different emotions have distinct spatial distributions that match expectations. We also perform temporal analysis using the capture time of the photos. Our spatio-temporal hotspot detection correctly identifies emerging concentrations of specific emotions and year-by-year analyses of select locations show there are strong temporal correlations between the predicted emotions and known events.

CCS Concepts

•**Computing methodologies** → **Scene understanding;** *Neural networks;* •**Human-centered computing** → *Geographic visualization;*

Keywords

Hotspot detection, emotion recognition, geotagged photos, spatio-temporal geographic analysis, deep learning

1. INTRODUCTION

Spatio-temporal hotspot detection is an important component of making cities smart, especially for tasks such as monitoring, early warning, resource allocation, and sustainable management. Hotspot analysis is typically conducted by mapping crime rates, monitoring disease outbreaks, locating traffic accidents, etc. In this paper, we focus instead on determining the emotional states of a city's inhabitants as conveyed through their photos as a step towards creating an affect-aware city.

Emotions play important roles in everyone's daily life. No matter what you do, you will have feelings associated with your activities. Services like Twitter, Facebook, Flickr, or Snapchat are great platforms for people to share their emotional states by posting words/pictures/videos. With geotagged and timestamped social multimedia, we can associate sentiment with geographical locations over time.

There exists work on detecting emotions from text such as in Twitter posts [12], but much less work on using image/video data. The reason is simple, words can deterministically express people's emotions, like "This is awesome!", "I feel blue", "So upset". However, predicting emotions in visual data is much more subtle and difficult. Luckily, the field of computer vision has made great advances recently in high-level image understanding thanks in large part to deep learning. With respect to our problem, large-scale visual datasets for emotion recognition [11, 14, 5] have been created allowing deep neural networks to be trained and achieve respectable performance on emotion recognition over the last five years. This has opened the opportunity for work such as ours to exploit these advances for performing spatio-temporal hotspot detection of public emotion using geo-referenced photos.

The major contributions of our work include: **(i)** We conduct the first investigation into sentiment hotspot detection in space and time via geotagged photos. **(ii)** The spatial hotspots for the different emotions have distinct spatial distributions and agree with expectations. **(iii)** Our temporal hotspot analysis is able to detect emerging concentrations of emotions. And, a year-by-year analysis of specific regions finds strong correlations between emotions and temporal events, such as between the level of joy and the success of the San Francisco Giants at AT&T park, and between the level of disgust and the increase in gentrification in the Mission residential neighborhood.

2. RELATED WORK

Our work is related to several lines of research.

Geo-Referenced Multimedia The exponential growth of publicly available geo-referenced multimedia has created a range of interesting opportunities to learn about our world. At the intersection of geographic information science and computer vision, large collections of geotagged photos have been used to map world phenomena [1], classify land use [15], geolocate photos [4], recognize and model landmarks [13], perform smart city and urban planning [10], etc. Although online photo collections represent a wealth of information, they present challenges due to how noisy and diverse they are. The challenges in using them for geographic discovery include inaccurate location information, uneven spatial distribution, and varying photographer intent. We are mindful of these and recognize they likely temper our results.

Deep Learning Deep learning has advanced a number of pattern recognition and machine learning areas including

SIGSPATIAL'16, October 31-November 03, 2016, Burlingame, CA, USA

DOI: http://dx.doi.org/10.1145/2996913.2996978

computer vision in which it debuted as deep convolutional neural networks (ConvNets) in 2012 [6]. Since then, researchers have applied deep ConvNets to a range of vision problems, obtaining state-of-the-art results. Key to ConvNets' performance is their ability to learn high-level or semantic features from the data as opposed to the hand-crafted low- to mid-level features traditionally used in image analysis. This level of image analysis, or understanding, is important for our task since detecting the emotion conveyed in an image is a high-level and abstract task.

Emotion Recognition Emotions represent higher intelligence and so being able to recognize them is key to artificial intelligence. For example, real-time emotion recognition during a customer service phone call can lead to a more satisfactory experience; analyzing a Twitter user's emotional state can help detect an emotional crisis; and, a chatting robot who is able to recognize emotions can have better interaction with users.

Emotions can be conveyed and therefore detected, at least in principle, in various multimedia sources such as text, images, and videos. We develop our own deep learning based system to detect the emotions conveyed in geotagged images. This then allows us to associate sentiment with place.

3. METHODOLOGY

We first describe our approach to detecting the emotion conveyed by an image. We then describe our spatial and spatio-temporal hotspot detection using the Gi* statistic [9] and Mann-Kendall test [8].

3.1 Emotion Recognition

As mentioned above, emotion can be conveyed by a number of multimedia sources such as text, audio, image, videos, etc. Visual emotion analysis is appealing since vision, as the richest sense, is arguably the most effective at conveying emotion. Existing work on visual emotion analysis can be classified into two approaches, dimensional models [7] and categorical models [11, 14]. We focus on categorical analysis using Ekman's six basic emotions [3]: *anger, disgust, fear, joy, sadness, and surprise.*

Our goal is using geotagged photos for sentiment hotspot detection. The foundation of our approach is assigning each photo one of the six emotions. We therefore design a per-image emotion classifier using ConvNets. This is motivated by the finding of You et al. [14] that ConvNets outperform traditional hand-crafted low-level features on most classes in a visual emotion analysis task. Specifically, we start with a VGG-16 network that has been pre-trained on ImageNet [2], and then fine-tune it using the Emotion6 dataset [11]. Once trained, our classifier achieves an average accuracy of 61.95%, which is reasonable and performs much better than random guess (which is 16.67%).

3.2 Spatial Hotspot Detection

Once we have labeled the geotagged images, we can map and start to investigate the spatial distribution of public sentiment. To simplify the analysis, we divide our study area, the city of San Francisco, into a 1000×1000 grid and assign each image to the closest bin center. The resulting quantization of the image locations does not affect our results since each bin measures less than approximately 12×14 meters which is finer than the scale of our analysis. All the spatial analysis below is based on the grid instead of the point locations of the photos.

Our data can now be considered a $1000 \times 1000 \times 6$ datacube in which the third dimension is the number of images labeled with a particular emotion. We normalize for the uneven spatial distribution of the images by computing the ratio of each emotion in each bin. That is, for each emotion, we compute a 1000×1000 grid where each bin is assigned

$$\text{ratio}_k^e = \frac{\text{number of photos in bin k of emotion e}}{\text{number of photos in bin k}}, \quad (1)$$

where k is the spatial index of the bin, and e is the emotion class. Each value in a bin indicates the percentage of a particular emotion evoked at the bin's location. Hence, for each location, the third dimension should sum to 1.

We use the Getis-Ord Gi* statistic [9] to find where high and low emotion ratios cluster spatially. Note that, for each emotion, only bins that contain photos are considered and nothing is computed for bins that do not have any photos.

3.3 Spatio-Temporal Hotspot Detection

Our geotagged photos have timestamps which enables us to perform temporal analysis. These timestamps indicate when the photo was taken. We temporally bin the photos at yearly intervals. Our photos span ten years so we now have a $1000 \times 1000 \times 10 \times 6$ datacube in which each bin is the ratio of images with a particular emotion to the total images for a particular location for a particular year. This now allows us to perform spatio-temporal hotspot detection.

Global Detection We perform spatio-temporal hotspot detection using the emerging hotspot analysis tool[1] in ArcGIS. We perform this analysis for each emotion separately. First, the Gi* statistic is computed spatially for each year. This is then followed by a Mann-Kendall test [8] to detect temporal trends at each spatial location. This test essentially looks for correlations between a spatial location's Gi* value and time. The emerging hotspot analysis tool classifies each spatial location into one of 17 categories: new hot (cold) spot, consecutive hot (cold) spot, intensifying hot (cold) spot, persistent hot (cold) spot, diminishing hot (cold) spot, sporadic hot (cold) spot, oscillating hot (cold) spot, historical hot (cold) spot, and no trend detected.

Local Temporal Analysis The emerging hotspot analysis tool identifies spatio-temporal hotspots but does not provide detailed information on the year-to-year changes. We therefore perform local analysis at a few locations with the goal of relating the changes to known temporal events. We explore a region's emotional trend over time by computing the emotion ratio for a bounding-box at yearly intervals:

$$\text{ratio}_y^e = \frac{\text{Number of photos locally of emotion e in year y}}{\text{Number of photos locally in year y}}. \quad (2)$$

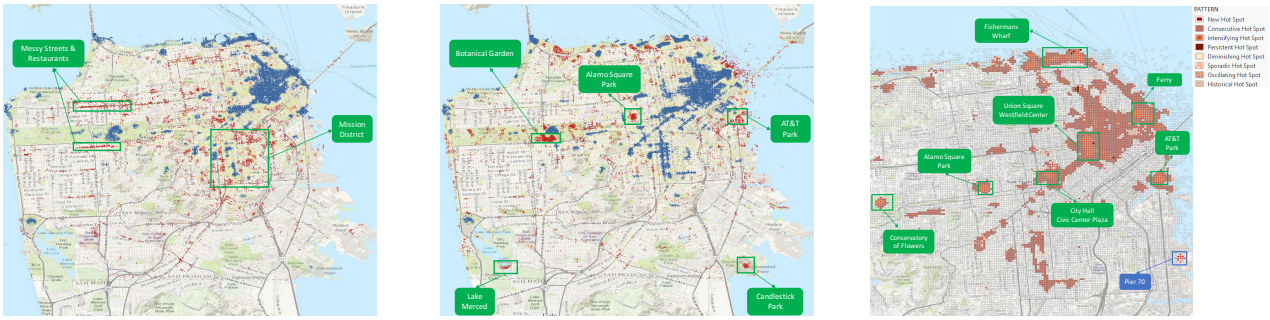
4. EXPERIMENTS AND RESULTS

We first describe the training and performance of our emotion classifier. We then present the results of the hotspot detection both in time and space.

4.1 Datasets

Emotion6 We use the Emotion6 [11] image dataset to fine-tune our emotion classifier. This dataset contains 1,980 images evenly divided into six emotion classes: anger, disgust, fear, joy, sadness, and surprise. The images were collected

[1]https://desktop.arcgis.com/en/arcmap/latest/tools/space-time-pattern-mining-toolbox/emerginghotspots.htm



(a) (b) (c)

Figure 1: Spatial hotspot detection: (a) disgust; (b) joy. Red, yellow, blue represent hot, not significant and cold spots, respectively. Spatio-temporal hotspot detection: (c) joy. See the text for more details.

from Flickr by using the class labels and synonyms as search terms. We randomly split the dataset into training and validation subsets in the ratio 8 : 2. All images are resized to 256×256 pixels for input to the classifier.

Geotagged Photos We download geotagged photos from Flickr for San Francisco city for the ten year period from 2006 to 2015. These are the images we label with our emotion classifier. The total number of images is around 1.9 million. However, some of the images are too dark/light, too small, or just a placeholder in Flickr, and so we perform a simple filtering step to remove these images. The dataset after filtering contains 1,753,903 images. The distribution by year and predicted emotion can be seen in figure 2.

Figure 2(a) conveys a sense of popularity of the Flickr platform over the ten year period. The number of uploaded photos reaches a peak of 259,741 in 2011 and then falls each year after that. This decline is interesting although we leave it to the reader to stipulate on its cause. Figure 2(b) shows the distribution of emotions as predicted by our classifier. There are more joy and sadness images than other emotions.

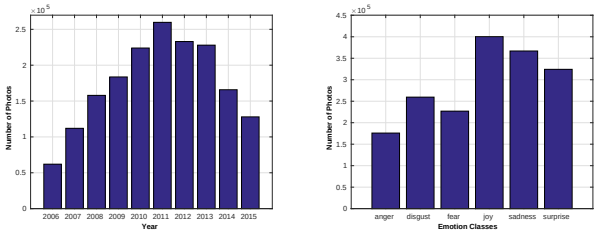


Figure 2: Number of geo-referenced photos in San Francisco area. Left: per year; right: per emotion.

4.2 Spatial Hotspot Analysis

We now present the results of our spatial hotspot detection. We use the optimized hotspot analysis tool[2] in ArcGIS to compute and visualize our results.

One of the challenges of our work is that there is no ground-truth for evaluation. Nonetheless, we make the following qualitative observations from the results in Fig. 1: **(i)** Different emotions have distinct spatial patterns (we only visualize the results of emotions joy and disgust for illustration). This indicates that our emotion classifier is detecting consistent signals in the geotagged photos. **(ii)** The detected hotspots make sense. For example, Fig. 1(b) shows that joy hotspots are detected at the San Francisco botanical garden, Alamo square park, AT&T park, Candlestick park, and the Fort Mason chapel. Further, these locations are detected as coldspots or not being significant for the other emotions. **(iii)** Some places are a mix of all emotions. These places, such as downtown San Francisco, have a wide variety of scenes which results in a relatively balanced distribution.

[2]http://desktop.arcgis.com/en/arcmap/10.3/tools/spatial-statistics-toolbox/optimized-hot-spot-analysis.htm

4.3 Spatio-Temporal Hotspot Detection

The goal here is to identify locations that are significiant in both space and time. We first conduct global detection and then perform temporal analysis for select locations.

Global Detection Fig. 1(c) shows the spatio-temporal hotspots detected for the north-east part of San Francisco. We make the following observations: **(i)** Pier 70 (blue box) is detected as a new hotspot which means that is a statistically significant hot spot for the final time step (2015) but has never been a statistically significant hot spot before. This makes sense since the Pier 70 buildings were recently renovated in 2014 to host large corporate parties, concert events, expositions, etc. **(ii)** Tourist destinations and public spaces are detected as intensifying hotspots which means they have been a statistically significant hot spot for ninety percent of the time-step intervals, including the final time step, and the intensity of clustering of high counts in each time step is increasing overall and that increase is statistically significant. The fact that these are intensifying and not just persistent hotspots is interesting. We postulate that it is due to the economic recovery that has occurred during the latter part of our time period which especially affects the tourist and leisure industry. **(iii)** Many locations are detected as consecutive hotspots which means there is a single uninterrupted run of statistically significant hot spot bins in the final time-step intervals but the location has never been a statistically significant hot spot prior to the final hot spot run and less than ninety percent of all time-steps are statistically significant hot spots. These locations also tend to be detected as cold spots or as having no significance in the spatial hopspot results in Fig. 1(b). Taken together, these results indicate that these locations have recently become hotspots. This again could be the result of the improved economy. It could also be the result of more photos being captured with GPS-enabled smartphones recently and thus having more accurate location information. This would make the photos more concentrated.

Local Analysis AT&T park shows up as a spatio-temporal hotspot with respect to joy. It is also a location whose sentiment one might expect to be correlated with the performance of the professional baseball team that plays there, the SF Giants. To investigate this, we calculate the joy ratio per

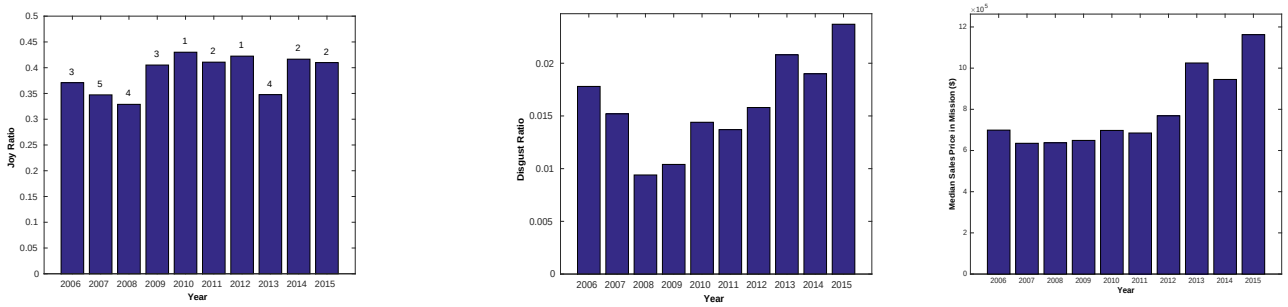


Figure 3: Observed temporal trends for select regions. Left: Joy ratio computed yearly for AT&T park. Shown above the bars are the end of season rankings of the SF Giants who play at the park. Notice the strong correlation. Middle: Disgust ratio computed yearly for the Mission neighborhood. Notice the steady increase since 2008 which is about when it started to become very popular with young professionals in the tech industry. Right: The average house price in the Mission neighborhood from a real estate website[3]. Notice the correlation with the disgust ratio.

year for a window centered on the park. These values are plotted in Fig. 3(a). Shown above each bar is the end-of-season ranking of the Giants for each year. The joy ratio and ranking are clearly correlated demonstrating that we are able to detect public sentiment from geotagged photos.

We also perform this local temporal analysis for another location, the Mission, for the emotion disgust. The Mission is one of the less expensive residential neighborhoods in San Francisco and is shown to exhibit a relatively large number of disgust spatial hotspots as shown Fig. 1(a) (this figure also delineates the neighborhood). We compute the per year disgust ratio in a window centered on the Mission and plot the results in Fig. 3(b). There is a clear increasing trend since 2008 which is about when the Mission started to become very popular with young professionals in the tech industry. These were not the traditional Mission residents and the detected increase in disgust could be a result of their reaction to the dirtiness, etc. of the streets. In fact, the yearly disgust ratio is strongly correlated with the average home price for the Mission, shown in Fig. 3(c)[3]. This increase in housing prices is likely also a result of the new demographic.

5. CONCLUSIONS

We conduct the first investigation into using geotagged social multimedia for spatio-temporal sentiment hotspot detection. We leverage deep ConvNets to develop an emotion classifier to predict the emotions conveyed in geotagged photos. This allows us to associate sentiment with place. We apply the Getis-Ord Gi* statistic to detect spatial hotspots, and show that different emotions have distinct spatial distributions that match expectations. We detect emerging concentrations of emotions through spatio-temporal hotspot detection and show that year-by-year analyses of select locations are correlated with known events.

6. ACKNOWLEDGMENTS

We gratefully acknowledge the support of NVIDIA Corporation through the donation of the Titan X GPU used in this work. This work was funded in part by a National Science Foundation CAREER grant, #IIS-1150115, and a seed grant from the Center for Information Technology in the Interest of Society (CITRIS). We would like to thank the UC Merced Spatial Analysis and Research Center (SpARC) for help with the hotspot analysis.

[3]http://www.trulia.com/real_estate/Mission-San_Francisco/1436/market-trends/

7. REFERENCES

[1] D. Crandall et al. Mapping the World's Photos. In *WWW*, 2009.

[2] J. Deng et al. ImageNet: A Large-Scale Hierarchical Image Database. In *CVPR*, 2009.

[3] P. Ekman et al. What Emotion Categories or Dimensions can Observers Judge from Facial Behavior. *Emotion in the Human Face*, 1982.

[4] J. Hays and A. A. Efros. IM2GPS: Estimating Geographic Information from a Single Image. In *CVPR*, 2008.

[5] Y.-G. Jiang et al. Speech Emotion Recognition Using Deep Neural Network and Extreme Learning Machine. In *AAAI*, 2014.

[6] A. Krizhevsky et al. ImageNet Classification with Deep Convolutional Neural Networks. In *NIPS*, 2012.

[7] X. Lu et al. On Shape and the Computability of Emotions. In *ACM MM*, 2012.

[8] H. B. Mann. Nonparametric Tests Against Trend. *Econometrica*, 1945.

[9] J. Ord and A. Getis. Local Spatial Autocorrelation Statistics: Distributional Issues and an Application. *Geographical Analysis*, 1995.

[10] S. Paldino et al. Urban Magnetism Through The Lens of Geo-tagged Photography. *arXiv preprint arXiv:1503.05502*, 2015.

[11] K.-C. Peng et al. A Mixed Bag of Emotions: Model, Predict, and Transfer Emotion Distributions. In *CVPR*, 2015.

[12] B. Resch et al. Citizen-Centric Urban Planning through Extracting Emotion Information from Twitter in an Interdisciplinary Space-Time-Linguistics Algorithm. *Urban Planning*, 2016.

[13] N. Snavely et al. Modeling the World from Internet Photo Collections. *IJCV*, 2008.

[14] Q. You et al. Building a Large Scale Dataset for Image Emotion Recognition: The Fine Print and The Benchmark. In *ACM MM*, 2016.

[15] Y. Zhu and S. Newsam. Land Use Classification Using Convolutional Neural Networks Applied to Ground-Level Images. In *ACM SIGSPATIAL*, 2015.