Published in the IEEE Applied Imagery Pattern Recognition Workshop, Washington, DC, October 13-15, 2015

Indexing Open Imagery to Create Tools to Fight Sex Trafficking

Abby Stylianou (a), Abigail Norling-Ruggles (b,c), Richard Souvenir (b,a), Robert Pless (a)
(a) Washington University in St Louis
(b) University of North Carolina at Charlotte
(c) Swarthmore College
corresponding author: abby@wustl.edu

Abstract

Images are important to fighting sex trafficking because they are: (a) used to advertise for sex services, (b) shared among criminal networks, and (c) connect a person in an image to the place where the image was taken. This work explores the ability to link images to indoor places in order to support the investigation and prosecution of criminal activity. We propose a framework which includes a database of open-source information available on the Internet, a crowd-sourcing approach to gathering additional images, and two baseline matching approaches. We concentrate on spatio-temporal indexing of hotel rooms, and to date have an index of more than 1.5 million geo-coded images. Our smart-phone app collects contextual information and metadata alongside images. On a test that included a database of 1800 images from 200 different hotels in St. Loouis, the correct hotel that matched a query images was found in the top 10 responses two-thirds of the time. We conclude with an analysis of the successes and limitations of our data set, our matching process, and suggestions for future research.

I. Introduction

Sex trafficking has become more prevalent as the Internet provides new avenues for the recruiting, advertisement and sales of sexual services [7]. In this work, we explore the opportunities of a highly connected world to build tools to fight sex trafficking. We particularly focus on understanding places where large scale image collection and analysis offers new investigative tools.

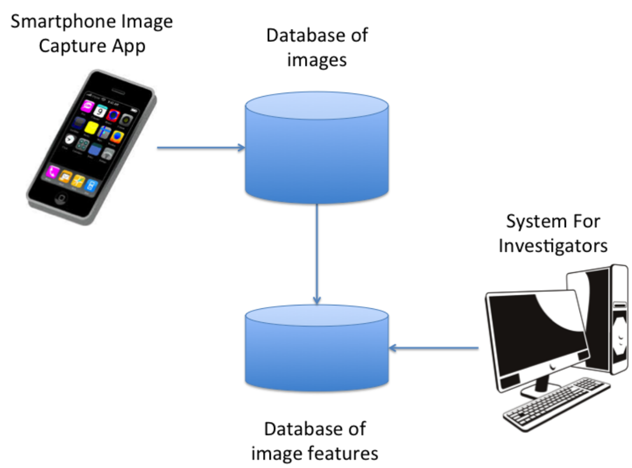


Fig. 1: A database of carefully annotated pictures of hotel rooms offers an important investigative tool. We explore frameworks that combine a crowd-sourcing mobile phone app that contributes to a database of images with a system that computes features from each image and offers an interface to find match query images to similar images in the database.

Images are a common way to advertise sex services. Images are interesting from an investigative standpoint because they connect the person in the image to the location where the image was taken. Therefore, they can help to characterize where a

978-1-4673-9558-8/15/31.00 ©2015$IEEE$

particular person was at different times. In the context of a sex trafficking investigation, this can be used to directly confirm that a person was in different states or countries. Among other things, this can change the set of laws under which a trafficker can be prosecuted.

This paper considers the potential to use the Internet to coordinate data capture efforts to fight sex trafficking. We aim to create a database of hotel room images that an investigator can use to understand the pictures they may acquire during a sex trafficking investigation. We build a dataset from publicly shared imagery on hotel booking sites, as well as from a smartphone app to crowd-sourcing the collection of pictures of hotel rooms. The crowd-sourcing option takes advantage of large scale trends in how people use social media; approximately 350 million photos are uploaded daily to Facebook. Tapping into this already common behavior creates the potential to rapidly create a relatively comprehensive, distributed, and continually updated resource that details the current appearance of hotel rooms worldwide. It also permits exploration into creating apps that encourage the acquisition of the most useful pictures for the matching process.

II. Background

We know of no prior efforts to explicitly match images to hotel rooms. Informal discussion with investigators reveal that the most commonly used technologies are manual searches through possible hotels in an area of interest, or using tools like Google "search by image" which returns images that are visually similar to a query. While Google does not advertise its proprietary searching method, there is a rich literature on content based image retrieval, and recent research on methods that scale up to large image database sizes [9]. In our paper, we explore baseline approaches based on SIFT features [8], and convolutional neural networks [6].

Our work seeks approaches to create Internet tools to fight sex-trafficking. It is widely understood that cyber-space markets are challenging places to coordinate efforts, because there is an asymmetry of incentives in cyber-space markets. Technology can be easily exploited by sex-traffickers to coordinate activities and advertising services. In contrast, the anti-trafficking efforts are sometimes hampered because the incentives for non-profit organizations often make them less likely to freely share resources as they struggle for recognition and funding for their efforts [1].

Open data has been used in trying to estimate the prevalence of trafficking [2] and to determine the effectiveness of US anti-trafficking funding projects [3]. Technology efforts have focussed on create search tools that index open data in different forms to create an interface that can be used for query and analysis [5], [12], [10], but to our knowledge there has not be any system that is explicitly focused on imagery.

III. Dataset Creation

In order to have the highest likelihood of finding a good feature match between a investigator's query image and the images in our dataset, our dataset should have as many images of as many rooms in as many hotels as possible. Additionally, it should have images from as many different times as possible. Hotels regularly renovate and change their internal appearance, meaning that photographs in our dataset may become outdated. These outdated images may still be valuable, however, in pinpointing the time frame in which an individual was trafficked (e.g., "This photograph was taken before the 2015 renovations, which means the person in the photograph was a minor at the time the advertisement was placed.").

We take two approaches to populating our dataset of hotel room images. First, we utilize already existing datasets of hotel room images used for marketing and travel sales. In particular, we keep track of the millions of images made available through Expedia's Affiliate Network API (http://developer.ean.com/) and create a reference in our database to the original data and its associated metadata. These photos, however, are often provided by the hotels themselves and may not present a comprehensive view of the hotel itself (e.g., only the nicest rooms from good angles in the best lighting). They may also not be updated following renovations. Both of those flaws would be problematic if these photographs were the only representations our dataset had of these hotels.

To supplement the images captured from existing datasets, we have created a smartphone crowd-

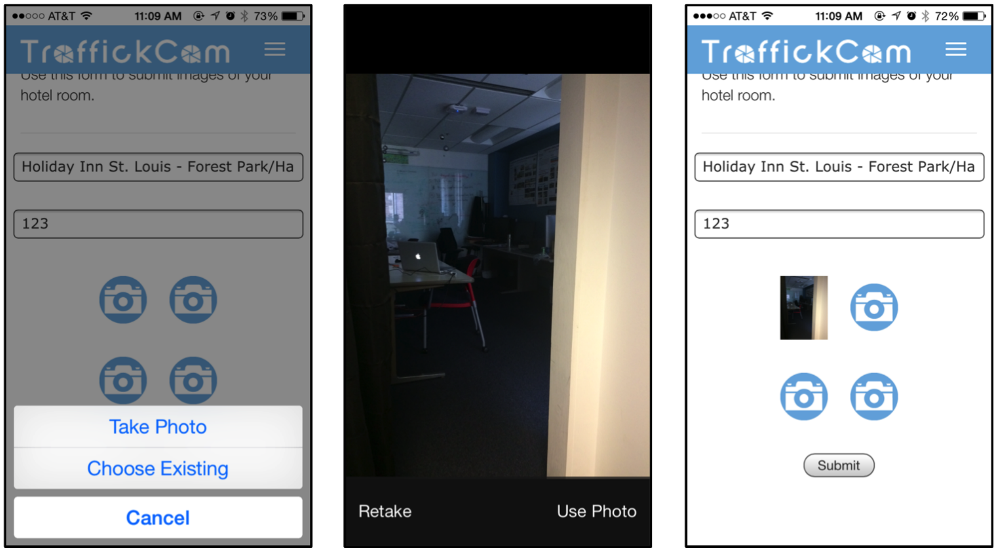


Fig. 2: Screenshots of a prototype smartphone app that allows anyone to contribute to the database.

sourcing application named TraffickCam, which allows travellers to upload their own photographs of a hotel room. This application is shown in Figure 2. Users are asked to provide minimal information regarding the photo – the name of the hotel they're staying in and their room number, along with images of the room.

At present, this dataset consists of 1,542,187 images taken at 145,846 hotels in the United States.

IV. Methods

We explore two baseline approaches for matching a query image to the dataset detailed in Section III. The first approach is based on SIFT feature matching, and the second approach is based on matching features extracted from the last layer (fc8) of the pre-trained Places network [14].

In the first approach, we extract SIFT features from every image in the dataset. Given a query image, we extract SIFT features. For each of those features, we find the n nearest neighbors in the set of features extracted from the database of images using the VLFeat MATLAB implementation of FLANN's KD-Tree Forests [11]. Each nearest neighbor match between a feature in the query image and a feature in a database image is a "vote" that the query image was taken in the same hotel as the database image. Votes are weighted by their ranking in the nearest neighbor match (e.g., the first nearest neighbor is weighted more heavily than the fifth nearest neighbor) to determine a list of candidate hotels where the query image might have been taken. This strategy is based on the approach to outdoor image localization in [13].

In the second approach, we extract feature representations learned from an existing deep convolutional neural network (CNN) architecture [6]. We use a publicly-available, pre-trained model, which we call *Places*, trained on the Places Database [14] for scene recognition from 205 categories (e.g., airplane cabin, hotel room, shed). In this CNN architecture, features are extracted from images in a layered, feed-forward manner. Initial layers of the architecture consist of convolutions, local response normalization, local pooling, dropout layers, and rectified linear (ReLU) activation units. The top layers of the network are four fully connected layers 'fc6', 'fc7', 'fc8', and the final output layer 'prob' that represents a categorical probability distribution. The dimensionality of these top layers in Places are 4096, 4096, 205, and 205 respectively. We perform feature extraction using Caffe [4], an open source deep learning framework.

V. Experimental Design and Results

We explore the accuracy of the methods suggested in the previous section with a experiment based on all hotels in St. Louis. We chose this scale of a test

Feature Set	Top 1	Top 10	Top 20
SIFT	**0.44**	**0.66**	**0.69**
Places (fc6)	0.32	0.63	**0.69**
Places (fc7)	0.26	0.54	0.65
Places (fc8)	0.14	0.44	0.52
Places (output)	0.04	0.25	0.31

TABLE I: Results with baseline feature matching methods. SIFT feature matching performance is better than features extracted from Places in identifying the correct hotel in the single most similar image (Top 1). SIFT features and features extracted from Places ('fc6') have similar performance in identifying the correct hotel in the Top 10 and Top 20 most similar images.

because in real use, investigations often have significant knowledge of where an image may be from, and because this scale makes the computational load small enough to easily test many different feature sets.

a) Dataset: In the St. Louis area, our database comprises 1800 images from about 200 hotels. We break this dataset into a database and a test set as follows. For every hotel that has at least 5 images, we choose one image as a test image and exclude it from the database. Exactly 100 hotels fit this criteria, creating a test set of 100 images. The 1700 remaining images were included in our database for this experiment.

b) Processing: For each query image, we follow the methodology detailed in Section IV, and compute the 20 nearest neighbors in the experimental database based on each of the feature type described in the previous section. For each query image, we find which hotel in which each of the 20 nearest neighbor images were captured, and report whether the correct hotel was in the top 1, top 5 and top 20 nearest neighbors.

c) Results: The results of this experiment are reported in Table I. SIFT feature matching generally has the best performance, identifying an image from the same hotel as the query image as the closest match 44% of the time. SIFT feature matching and matching using the feature extracted from Places layer 'fc6' have similar performance when identifying the correct hotel in the top 10 and top 20 closest matches. The places 'output' layer has generally poor performance. We show example results for SIFT feature matching in Figure 3.

VI. Discussion and Future Work

This paper details the beginning of a project to find existing Internet imagery and crowd-source the collection of additional imagery of hotel rooms. The dataset creates a resource that can be used in investigations of sex-trafficking because it provides possible locations where photographs of sex-trafficking victims were taken. Our initial results are promising on an experimental dataset that includes thousands of images from all the hotels in a city. Qualitatively, this is a test on a scale that may itself be useful (if the investigation already knows to focus on a city). Also qualitatively, a search that half the times gives exactly the correct hotel is likely to be a useful tool.

Quantitatively, the study suggests that SIFT features designed to find exact matches are competitive with state of the art Convolutional Neural Network approaches. There are many reasonable approaches to attempt to improve the performance of both approaches, and we look forward to understand which are these are best when the query is tested against a global scale database instead of one city.

This study had several limitations. First, the query images in our test were drawn from our database of hotel rooms. While they were not used in the training of the CNN, they also did not have any people, and did not have the variations in lighting, pose, etc. that might be part of real query images. Second, real query images, especially those used in sex advertising, are not focused on showing the hotel room; it is likely that only the edges of such query images are helpful to detecting the hotel room. For an approach based on SIFT features, it is relatively easy to discard features that are not from parts of the image that show features of the room. For a CNN based approach, the image features computed (in layers fc6, fc7) are affected by the entire image, and it is less clear how to encourage the search to focus on the background. This is an interesting area of future work.

References

[1] D. Blair. It takes a movement to defeat a market: Tech-enabled collaboration to combat human trafficking. In *APSA 2014 Annual Meeting Paper*, 2014.

[2] L. Fedina. A systematic review and methodological critique of human trafficking prevalence studies. In *Society for Social Work and Research 19th Annual Conference: The Social and Behavioral Importance of Increased Longevity*. Sswr, 2015.

[3] K. Fedorschak, S. Kandala, K. C. Desouza, and R. Krishnamurthy. Data analytics and human trafficking. In *Advancing the Impact of Design Science: Moving from Theory to Practice*, pages 69–84. Springer, 2014.

[4] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell. Caffe: Convolutional architecture for fast feature embedding. *arXiv preprint arXiv:1408.5093*, 2014.

[5] C. A. Knoblock and P. Szekely. A scalable architecture for extracting, aligning, linking, and visualizing multi-int data. In *SPIE Sensing Technology+ Applications*, pages 949907–949907. International Society for Optics and Photonics, 2015.

[6] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*, 2012.

[7] E. I. Kunze. Sex trafficking via the internet: How international agreements address the problem and fail to go far enough. *J. High Tech. L.*, 10:241, 2009.

[8] D. G. Lowe. Object recognition from local scale-invariant features. In *Computer vision, 1999. The proceedings of the seventh IEEE international conference on*, volume 2, pages 1150–1157. Ieee, 1999.

[9] D. Moise, D. Shestakov, G. Gudmundsson, and L. Amsaleg. Indexing and searching 100m images with map-reduce. In *Proceedings of the 3rd ACM conference on International conference on multimedia retrieval*, pages 17–24. ACM, 2013.

[10] D. R. Silva, A. Philpot, A. Sundararajan, N. M. Bryan, and E. Hovy. Data integration from open internet sources and network detection to combat underage sex trafficking. In *Proceedings of the 15th Annual International Conference on Digital Government Research*, dg.o '14, pages 86–90, New York, NY, USA, 2014. ACM.

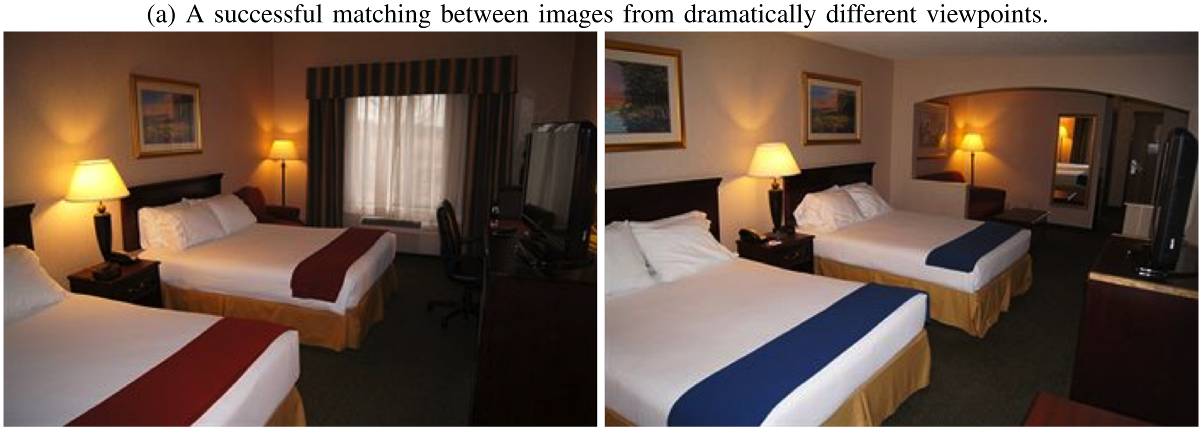
[11] A. Vedaldi and B. Fulkerson. VLFeat: An open and portable library of computer vision algorithms. http://www.vlfeat.org/, 2008.

[12] H. Wang, C. Cai, A. Philpot, M. Latonero, E. H. Hovy, and D. Metzler. Data integration from open internet sources to combat sex trafficking of minors. In *Proceedings of the 13th Annual International Conference on Digital Government Research*, pages 246–252. ACM, 2012.

[13] A. R. Zamir and M. Shah. Accurate image localization based on google maps street view. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2010.

[14] B. Zhou, A. Lapedriza, J. Xiao, A. Torralba, and A. Oliva. Learning Deep Features for Scene Recognition using Places Database. In *Advances in Neural Information Processing Systems*, 2014.

(a) A successful matching between images from dramatically different viewpoints.



(b) A successful matching that demonstrates the limitations of our current dataset. These two images are more visually similar than we would ever expect in real world query data.

(c) A failed matching, where SIFT feature matching found visually similar features in the furniture in hotel rooms in two different hotels.

Fig. 3: The left column shows query images, and the right image shows the image which was found to be the closest match using SIFT features and the matching pipeline described in Section IV. The top two rows show correctly matched pairs, where the query image and result image were taken in the same hotel. The bottom row shows an incorrectly matched pair.