Zebra Crossing Detection from Aerial Imagery Across Countries

Daniel Koester, Björn Lunt, and Rainer Stiefelhagen

Karlsruhe Institute of Technology, Karlsruhe, Germany
{daniel.koester,bjoern.lunt,rainer.stiefelhagen}@kit.edu

Abstract. We propose a data driven approach to detect zebra crossings in aerial imagery. The system automatically learns an appearance model from available geospatial data for an examined region. HOG as well as LBPH features, in combination with a SVM, yield state of the art detection results on different datasets. We also use this classifier across datasets obtained from different countries, to facilitate detections without requiring any additional geospatial data for that specific region. The approach is capable of searching for further, yet uncharted, zebra crossings in the data. Information gained from this work can be used to generate new zebra crossing databases or improve existing ones, which are especially useful in navigational assistance systems for visually impaired people. We show the usefulness of the proposed approach and plan to use this research as part of a larger guidance system.

Keywords: Aerial Imagery · Street Crossing · Visually Impaired

1 Introduction

Visually impaired and blind people face a multitude of challenges in many aspects of their daily lives. Among those, outdoor as well as indoor navigation remains a great challenge due to the complexity of their surroundings and interactions with moving obstacles, such as other pedestrians or vehicles. Much needed visual cues are constantly utilized by unimpaired people, to avoid accidents and to not bump into one of the numerous obstacles, and are thus taken for granted. This effect is further amplified by an urban element design that neglects visually impaired peoples' specific needs. Sidewalks are often not equipped with tactile pavings and items such as trash cans, lamp poles, or safety posts are often placed in spatially confusing or free space limiting locations, severely impeding movements by visually impaired people. Street crossings, as well as pedestrian traffic lights, often do not include much needed accessibility features and the general street layout can be very confusing to the visually impaired, e.g., Y-junctions, complicated round-abouts, or temporary construction sites. With respect to accessibility, zebra crossings are amongst the worst offenders, often providing no locational hints at all and only very limited tactile feedback while crossing the street. Furthermore, they provide less certainty and peace of mind compared to formal crossings [1].

K. Miesenberger et al. (Eds.): ICCHP 2016, Part II, LNCS 9759, pp. 1–8, 2016.
DOI: 10.1007/978-3-319-41267-2_5

Today's ubiquitous *Global Navigation Satellite Systems* allow visually impaired people to regain at least some confidence in urban navigation situations. While these can greatly increase mobility, they only provide very high-level navigational information, but can neither warn of obstacles nor supply fine level guidance to the user. Thus, to this day, visually impaired people rely traditionally most on their white cane or a guide dog for these types of information. Furthermore, important geospatial information databases for visually impaired people are often under-utilized or non-existing. Our work therefore tries to improve already existing geospatial databases, such as *OpenStreetMap*[1] (OSM). It's data quality varies a lot as it is collected and edited by volunteers all around the world. Additionally, the specific details of a street crossing, e.g., whether it is equipped with pedestrian traffic lights, is often missing or even incorrect. We plan to use such weekly annotated data in order to create an assistive system that helps visually impaired people cross the street more safely. This work focuses especially on using nearby zebra crossings, as they provide much more safety to visually impaired people compared to just plain crossing the street. Very few specialized navigational assistance systems use such available databases already, e.g., *Trekker Breeze* or *Braille Note GPS*[2], however, these often contain only points of interest or require pre-recorded routes, thus limiting their usefulness. We believe that an increased number and improved quality of available geospatial databases for visually impaired people will aide their usefulness and greatly increase utilization and integration in the near future. Furthermore, such data could be helpful for cars, as they could be made aware of an oncoming zebra crossing and react accordingly.

This work leverages available geospatial data in order to create a training corpus for a data driven approach. We acquire aerial imagery from different sources, such as *Aerowest*[3] or *Google Maps*[4]. The aerial imagery is then pre-processed using information about road locations and directions that are extracted from OSM meta data. This process greatly reduces the considered search space in the aerial imagery, strongly depending on the examined region. We then extract a combination of *Histogram of Oriented Gradients* (HOG) [2] as well as *Local Binary Pattern Histograms* (LBPH) [3] descriptors and train a classical *Support Vector Machine* (SVM) [4] for classification. We also compare it to a line search based approach, using regional data that is similar in appearance, as well as a classifier trained on our own dataset. Furthermore, our resulting classifier is also used to try to detect yet uncharted zebra crossings in the aerial imagery, which were then verified by hand. Our system is widely usable, as it can automatically learn the local appearance of zebra crossing patterns in different countries, which often have different regulations and therefore zebra crossings differ in visual appearance. Finally, our system is capable of creating such a geospatial database for regions where it doesn't yet exist.

[1] http://www.openstreetmap.org
[2] http://www.humanware.com
[3] http://www.aerowest.de
[4] http://maps.google.com

2 Related Work

Using aerial imagery of urban areas to create geospatial data has wide applications, e.g., city planning or road detection and analysis [5], and has been studied for a long time [6]. Adding geospatial data to map services [7], specifically the detection of street-level accessibility problems [8] or points of interest for visually impaired people [9,10], has seen lots of interest in recent years.

Detection of zebra crossings from street level scenarios for visually impaired people has already been researched for quite some time. Usually, line detection algorithms are used in combination with other features. Se [11] combines these with intensity variation as well as pose information, to distinguish between zebra crossings and very similar looking staircases. Ahmetovic et al. [12,13] use the position of the horizon obtained through a mobile phone's accelerometer to reduce the search space in combination with their line detection based *ZebraRecognizer* library. Coughlan et al. [14] and Ivanchenko et al. [15,16] also largely base their work on the same approach, and provide user centered features on top of it.

Zebra crossing detection from aerial imagery was, just recently, first proposed by Ahmetovic et al. [17]. The authors use aerial imagery from *Google Maps* and download only partitions that actually contain road surface. They achieve this by calculating the distance of the partition's center to the next road using the *Google Maps Javascript API* and verify zebra crossing candidates with *Google Street View* panoramas. Their approach uses modified line search algorithms for both detection stages, in the aerial imagery as well as acquired panoramas. This two-step approach is evaluated on a 1.6km^2 rectangle in a San Francisco neighbourhood. We compare their approach to ours using the same region downloaded from *Google Maps* and achieve competitive results.

3 Methodology

The proposed data driven algorithm is based on available geospatial data and aerial imagery, by providers such as *OpenStreetMap*, *Google Static Maps*, or *AeroWest*. Such aerial imagery providers allow anyone to retrieve images for specific regions – usually only single and rate limited tile-based requests of varying size – that are centered around a user defined longitude and latitude coordinate date. Using OSM meta data from its *Overpass API*, we acquire the precise positions of road connection points, referred to as nodes, as well as some already labeled zebra crossings. We clean the data to save on computational time, e.g., neglect parking areas, private property, and factory sites, but it is not required. With knowledge gained from this meta data, we download a region of interest's connected and neighbouring parts, focussing on the road surface. We move from node to node in a linear and interpolated fashion, which represent the estimated street locations and acquire the necessary aerial imagery for further processing. Although the meta data's nodes are not always perfectly aligned to the actual road, their observed quality has always proven good enough to capture most, if not all, of the desired road surface.

A sliding window approach that follows the road structure given by the OSM meta data's nodes in the previous step then scans the obtained imagery and saves each window for further processing. This process also aligns it with respect to the road center. This crucial step removes almost any rotational variance of observed zebra crossings, and also drastically reduces the search space and simplifies the classification and detection tasks.

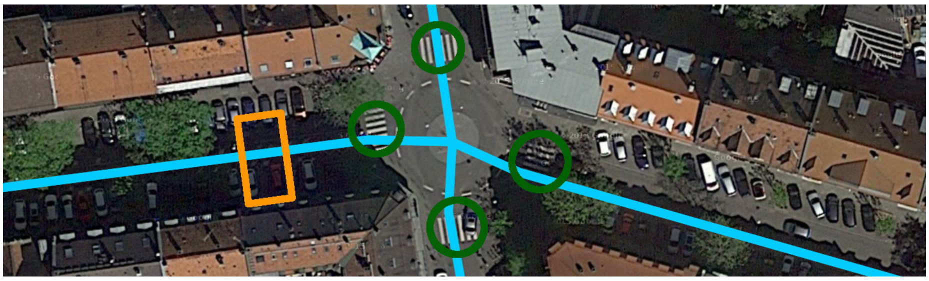

Fig. 1. An example of combined aerial imagery and meta data showing a small roundabout in a German city. Center lines (blue/bright) often deviate from the ideal position. Zebra crossings (green/dark circles) might be occluded by cars or subject to shadows. Our algorithm uses a sliding window (orange/bright rectangle) along the center lines. Image source: Google.

Fig. 1 shows an example of the combined data with the road center lines, labeled zebra crossings and the sliding window. We use a HOG implementation created by Felzenszwalb et al. [18]. While zebra-crossings are often occluded by shadows and trees or have deteriorated markings, HOG block normalization helps to compensate for such variances. HOG features are suited to this task, because they describe local intensity gradients, i.e., edge directions, which appear quite regular on zebra crossings from an aerial viewpoint. We also compute LBP features on the image window and use their normalized histogram as an input vector for the SVM. Additionally, we also train SVMs with a combined HOG and LBPH feature vector, but the improvements in accuracy and average precision are negligible. We then extract the known zebra crossing locations from the acquired scan windows for a processed region and use HOG and LBPH features with varying parameters to train a SVM. We train linear – much faster to train and test – and *radial based function* (RBF) – yield slightly better results – SVM kernels and compare their results. In a last step, the classification results are clustered using direction information from the sliding window movement as well as distance measures, and these yield our final zebra crossing detections.

After training, we rescan the aerial imagery, or evaluate the classifier on different data, and use the trained SVM to detect zebra crossings, often yielding zebra crossings not present in the meta data. Furthermore, it greatly reduces false positive detections on similar looking structures, e.g., roof tops or staircases, as only actual road surface can be considered by the algorithm.

4 Evaluation

We use aerial ortho-imagery, i.e., pixels correlated to precise latitude and longitude locations. It was collected from various urban and rural regions in Germany and sampled from ~10km of road surface, which contain 3119 zebra-crossings. Any encountered zebra crossings are extracted without regards to the observed image quality. Hence, these contain a great deal of variance: Occlusion from vehicles or trees, deterioration of markings, shadows from nearby trees and buildings, general illumination changes from different daytimes as well as even varying seasons. The resolution varies from 5 to 10cm per pixel and has to be considered in any pixel-to-coordinate translation.

Table 1. Evaluation results on our own dataset (3119 zebra crossings), where *"-lin"* denotes a linear and *"-RBF"* a *radial basis function* SVM kernel, "HOG^{30x30}" is the block size, "LBP$^{17/10}$" are the *radius/neighbour* variation. Numbers given in %.

Method	Precision	Recall	Accuracy	Avg.-Prec.
HOG^{30x30}-lin	74.8	93.1	92.4	94.43
HOG^{20x20}-RBF	95.2	96.2	**98.9**	97.99
LBP$^{17/10}$-lin	99.4	**97.4**	98.4	**99.56**
LBP$^{17/10}$-RBF	**99.7**	97.0	98.3	**99.56**

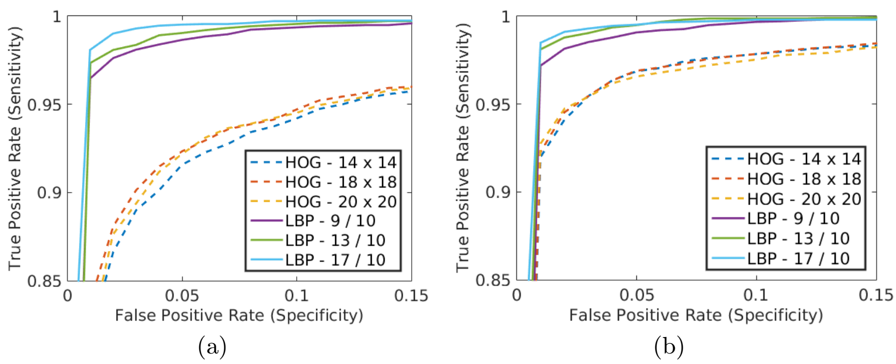

Fig. 2. ROC curves for the 5-fold-cross validation on our own dataset: (a) uses a linear SVM kernel, while (b) uses an RBF kernel. Both figures show performance for various HOG block sizes, while LBP parameters are *radius/neighbour* variations.

We train our algorithm only on a subset of the available data, due to the large number of negative samples, and use data augmentation, i.e., we mirror all positive samples horizontally, vertically, and in combination. Our train and test set for the 5-fold-cross validation then consists of as many positive as randomly chosen negative samples, i.e., ~12500 each, we will consider hard negative mining instead in the future. This augmented and reduced dataset is then used to test different parameter configurations, a process that would otherwise be computationally infeasible when including all negative samples. After testing and subsequently choosing different parameter configurations for LBPH and HOG, as well as fine tuning the SVM's parameters, we compare the classification results in Table 1 and show that our algorithm achieves close to perfect performance. Furthermore, we show *ROC* curves for some of the different tested HOG and LBPH parameters in Fig. 2 and observe that performance varies only slightly, while LBPH features perform constantly better than HOG features.

Table 2. Comparison to Ahmetovic et al. [17] (141 zebra crossings inside a San Francisco rectangle). We report their satellite-data only (*"-SAT"*), street-view refinement (*"-SV"*) and combined results as well as our algorithm trained on surrounding data and a classifier based on our own dataset (*"-PRE"*). All numbers given in % and optimized for recall.

	Precision	Recall
[17]-SAT	68.8	97.2
[17]-SV	97.2	97.8
[17]-SAT+SV	97.2	95.0
Ours	96.2	95.7
Ours-PRE	98.9	38.4

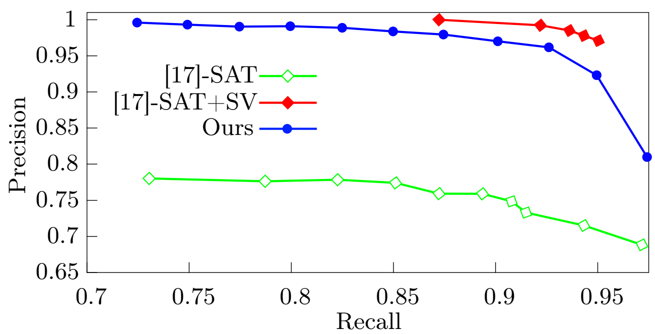

Fig. 3. ROC curve for the San Francisco dataset showing Ahmetovic et al.'s [17] performance of individual steps versus our classifier. Original figure taken from [17], updated with our (discretized) data and modified for improved readibility.

In addition to the cross validation, Table 2 and Fig. 3 compare our algorithm to Ahmetovic et al. [17], which tuned their algorithm for maximum recall to minimize the chance of missing any zebra crossings. To make a fair comparison between the different approaches, we download the same San Francisco rectangular region from *Google Maps*. In order to train a SVM with similar data, we manually search for ∼250 zebra crossings within the direct neighbourhoods as well as cities close by, as OSM meta data for zebra crossings was not available, a state which we hope to eventually change with our system. Furthermore, we use our pre-trained SVM classifier from our own German dataset for detection purposes and compare the results accordingly.

For ground truth, we had to manually label the rectangle's zebra crossings and count 184 zebra crossings overall: 122 are "*continental crossings*" (typical crossing type found in the U.S. [17]) and 62 that also contain "*transverse markings*" (two additional white perpendicular lines, resulting in a *ladder* like appearance). These numbers differ slightly from Ahmetovic et al. [17], mainly due to us considering *ladder* like zebra crossings as ground truth. We also ignore the 110 plain "*transverse markings*" found on many of the other street crossings, similar to the work we compare to. Detecting these would be a much harder task, as they only consists of 2 parallel lines.

Our data driven algorithm achieves competitive results and even our *across-countries* (Ours-PRE) classifier performs reasonable. It's results might be used further to train a new classifier using it's top and bottom k-elements instead of having to manually search for training data, as we had to. Such an approach would allow to generate missing geospatial databases for almost any region.

5 Conclusion

We demonstrate a system to improve the general availability and quality of geospatial data, i.e., zebra crossing locations, to be used in navigation and guidance applications for visually impaired people. Our proposed detection algorithm, which learns automatically from provided data, consists of HOG and LBPH features and uses a SVM for classification, achieves state of the art results on our own dataset as well as competitive results on those proposed by other authors. Moreover, we show that our algorithm can perform reasonably well when trained and tested on different data sources, providing a decent cross-dataset performance. Such cross-dataset results might then be used as an intermediate step in order to generate training data for a new generation of classifiers. This process even allows to involve a human to ensure a sufficient training set quality.

It is our belief that the availability of accurate geospatial databases will help foster their integration into products for visually impaired people as well as traditional or autonomous cars, and thus increase the safety on the road. We plan to further investigate our approach, to improve its performance as well as eventually use it in a more complete guidance system for visually impaired people, which not only navigates to the next available street crossing but also helps to safely cross the road. Future research will therefore involve the creation of such a guidance system that detects zebra crossings from a street perspective, recommends the next zebra crossing location using a geospatial database and guides the user towards, as well as safely over, the zebra crossing itself. Finally, studies with visually impaired people will be conducted, as part of a research effort to create a complete guidance solution to be used in urban or rural locations.

References

1. Matthews, B., Hibberd, D., Carsten, O.: Road and street crossings for blind and partially sighted people: the importance of being certain (2014)
2. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 886–893 (2005)
3. Ojala, T., Pietikäinen, M., Mäenpää, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans. Pattern Anal. Mach. Intell. **24**(7), 971–987 (2002)
4. Cortes, C., Vapnik, V.: Support-vector networks. Mach. Learn. **20**(3), 273–297 (1995)
5. Hintz, S., Baumgartner, A.: Automatic extraction of urban road networks from multi-view aerial imagery. ISPRS J. Photogrammetry Remote Sens. **58**(1–2), 83–98 (2003)
6. Quam, L.H.: Road tracking and anomaly detection in aerial imagery. In: ARPA Spring 1978 Image Understanding Workshop (1978)
7. Goodchild, M.F.: Citizens as sensors: the world of volunteered geography. GeoJ. **69**(4), 211–221 (2007)
8. Hara, K., Le, V., Froehlich, J.: Combining crowdsourcing and google street view to identify street-level accessibility problems. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 631–640 (2013)
9. Prasain, S.: Stopfinder: improving the experience of blind public transit riders with crowdsourcing. In: Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility, pp. 323–324 (2011)
10. Rice, M., Aburizaiza, A., Jacobson, D., Shore, B., Paez, F.: Supporting accessibility for blind and vision-impaired people with a localized gazetteer and open source geotechnology. Trans. GIS **16**(2), 177–190 (2012)
11. Se, S.: Zebra-crossing detection for the partially sighted. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 211–217 (2000)
12. Ahmetovic, D., Bernareggi, C., Mascetti, S.: Zebralocalizer: identification and localization of pedestrian crossings. In: Proceedings of the International Conference on Human Computer Interaction with Mobile Devices and Services, pp. 275–286 (2011)
13. Ahmetovic, D., Bernareggi, C., Gerino, A., Mascetti, S.: ZebraRecognizer: efficient and precise localization of pedestrian crossings. In: 22nd International Conference on Pattern Recognition (ICPR), pp. 2566–2571 (2014)
14. Coughlan, J., Shen, H.: A fast algorithm for finding crosswalks using figure-ground segmentation. In: Proceedings of 2nd Workshop on Applications of Computer Vision, in Conjunction with ECCV (2006)
15. Ivanchenko, V., Coughlan, J., Shen, H.: Detecting and locating crosswalks using a camera phone. In: IEEE Conference on Computer Vision and Pattern Recognition Workshops (2008)
16. Ivanchenko, V., Coughlan, J.M., Shen, H.: Crosswatch: a camera phone system for orienting visually impaired pedestrians at traffic intersections. In: Miesenberger, K., Klaus, J., Zagler, W.L., Karshmer, A.I. (eds.) ICCHP 2008. LNCS, vol. 5105, pp. 1122–1128. Springer, Heidelberg (2008)
17. Ahmetovic, D., Manduchi, R., Coughlan, J., Mascetti, S.: Zebra crossing spotter: automatic population of spatial databases for increased safety of blind travelers. In: Proceedings of the 17th International ACM SIGACCESS Conference on Computers and Accessibility, pp. 251–258 (2015)
18. Felzenszwalb, P.F., Grishick, R.B., McAllester, D., Ramanan, D.: Object detection with discriminatively trained part based models. IEEE Trans. Pattern Anal. Mach. Intell. **32**(9), 1627–1645 (2009)