

A Paradigm Shift: Detecting Human Rights Violations Through Web Images

Grigorios Kalliatakis, Shoaib Ehsan, and Klaus D. McDonald-Maier

School of Computer Science and Electronic Engineering, University of Essex, UK
{gkallia, sehsan, kdm}@essex.ac.uk

Abstract. The growing presence of devices carrying digital cameras, such as mobile phones and tablets, combined with ever improving internet networks have enabled ordinary citizens, victims of human rights abuse, and participants in armed conflicts, protests, and disaster situations to capture and share via social media networks images and videos of specific events. This paper discusses the potential of images in human rights context including the opportunities and challenges they present. This study demonstrates that real-world images have the capacity to contribute complementary data to operational human rights monitoring efforts when combined with novel computer vision approaches. The analysis is concluded by arguing that if images are to be used effectively to detect and identify human rights violations by rights advocates, greater attention to gathering task-specific visual concepts from large-scale web images is required.

Keywords: image understanding, human rights violations recognition, human rights monitoring

1 Introduction

Digital technologies provide new openings for human rights advocates, including non-governmental organizations (NGOs) such as Amnesty International and Human Rights Watch, to identify, expose, verify and document human rights violations [11]. These emerging technologies are likely to complement conventional interview-based fact-finding approaches in order to effectively verify abuses. In this paper, we use the term and concept of ‘human rights violations’ in keeping up with the ways it has being adopted by human rights advocates, that is, to refer to actions executed by state or non-state actors that breach any part of those rights that protect individuals and groups from behaviours that intervene with fundamental freedoms and human dignity [12].

Digital images hold immense potential for monitoring emerging human rights violations and emphasising on specific events that might otherwise elude attention. Organizations concerned with human rights advocacy are increasingly using digital images as a tool for improving the exposure of human rights and international humanitarian law violations that may otherwise be impossible. However, these organizations continue to depend mainly on qualified imagery analysts to

manually identify human rights violations events from the extensive amount of data. Because of the high expense involved concerning these analysts [8], very few organizations are able to conduct monitoring campaigns over large-scale data.

Until now, all human rights event detection efforts are implemented by utilising tools such as remote sensing platforms and social media. The aim of this paper is to discuss the value of digital images within the context of detecting human rights violations, in order to analyse major opportunities and difficulties for their growing and prolonged use in supporting human rights. Our research lens is focused on establishing a new approach for human rights advocates to monitor different forms of human rights violations without requiring the use of trained imagery analysts.

The remainder of the paper is organised as follows. Section 2 examines prior works related to human rights monitoring tasks throughout remote sensing and social media. Section 3 examines the value of real-world images for detecting human rights violations. Also in the same section the difficulties of collecting real-world images from the web are thoroughly discussed. Finally, conclusions and future directions are given in section 4.

2 Background

Human rights monitoring is an extensive term outlining the collection, verification, and use of information to address human rights problems. The need of regularly detecting phenomena associated with some violations has led organizations to conduct various human rights monitoring campaigns. Up to now, the process of detecting international law violations engages solely two main mechanisms, *remote sensing* and *social media* alongside eyewitness reports. This section will briefly review those two research directions emphasizing on their shortcomings.

2.1 Remote Sensing

Remote sensing is considered a mechanism capable of improving the process of detecting human rights violations and is nowadays performed by platforms such as helicopters and unmanned aerial vehicles where people are at risk of suffering from such offences. Organizations that monitor human rights violations have increasingly followed a typical sequence, as described in [8], of remote sensing. First the violation which they are seeking to detect needs to be determined. After that, a remotely sensed phenomenon associated with that particular violation is chosen and finally a suitable sensor that will allow the detection of this phenomenon is established. Satellites can be utilised to detect phenomena such as extensive changes to villages [9], identifying mass graves and large-scale destruction of civilian properties. However, most phenomena currently being exploited in human rights monitoring campaigns, demand a high-resolution sensor. Another drawback can be found when dealing with specific atrocities, such as genocide, where a systematic and a time-series analysis is required for the evidence. The potential for satellite images to be used in human rights contexts

has not been fully realised yet, mainly because of the required cost and limited resources.

2.2 Social Media

Another mechanism suited for information-gathering tasks by human rights organizations is social media. Such organizations, including NGOs, are currently analysing data from social media, such as Twitter and Facebook, to a greater extent in order to recognise, verify and document different human rights abuses. Social media are capable of circulating information earlier and more comfortably compared to traditional media. Furthermore, social media manage to provide evidence of different types of human rights violations that can avoid being censored by restrictive governments. However, manually identifying human rights violations events from such an extensive amount of data poses serious technical challenges for human rights NGOs. In order to address this challenge, [2] proposed an approach for automated, large-scale analysis of human rights related events. They model social media (such as Twitter) data as a heterogeneous network with multiple different nodes, relationships and features by employing a Non-Parametric Heterogeneous Graph Scan (NPHGS), in order to achieve human rights events identification and characterization from the emerging patterns.

3 Images for Detecting Human Rights Violations

Images have grown into omnipresent on the Internet, something that has inspired the already highly motivated process of analysing their semantic content for various applications. From this perspective, digital images have growing usage among the human rights sector. Such organizations are progressively attempting to exploit the capacity of digital images as a tool for identifying, exposing and verifying different forms of human rights and international humanitarian law abuses. As discussed in Section 2.1 there are several limitations for human rights-focused organisations when utilising remote sensing such as: 1) high cost for acquiring high-resolution images; 2) restricted access to valuable resources; 3) limited efficacy of satellite images which leads only to large-scale human rights violations detection; 4) the obligation for qualified imagery analysts and 5) specific sensor characteristics as prerequisites. Such difficulties can be overcome by utilising a combination of public available digital images and a novel context-based vision system capable of detecting human rights violations. The most promising aspect of real-world digital images compared to satellite images is that they can be effortlessly captured by using everyday devices such as mobile phones. Consequently, the cost for acquiring such images is practically negligible while their number can rapidly expand without the need of commercial providers.

3.1 Digital Images as Legal Evidence

Certainly, the efficiency of publicly available real-world images varies according to purpose and the nature of the images collected. Gathering such evidence can

IV



Fig. 1. Object recognition results for two images. Both images contain a gun but it is evident that only the left image can be utilised for the task of recognising human rights violations.

serve not only for advocacy purposes but potentially for evidence in legal cases as well. Evidence is anything that can provide information about an incident being investigated as stated in [10] and may come from many different resources like physical, testimony, documents and imagery. Our future research will focus on real-world images in order to exploit the high-standard characteristics of legal evidence to human rights monitoring process. Typically there are three main questions set by every human rights monitoring mechanism: *what*, *who* and *how*. Examples of such crimes include the following: 1) Military forces torturing a civilian; 2) Bulldozers unlawfully destroying homes; 3) Children in military camps being trained for warfare and 4) Police placing a suspect in an illegal chokehold. Also documenting the who and how, often referred to as *linkage evidence*, can provide important clues to investigators. For example 1) the face of a soldier as he repeatedly beats a civilian; 2) the face of the bulldozer operator as he destroys homes and 3) the badge number and name plate of an officer who is illegally holding a suspect.

Object recognition in computer vision is the task of finding and identifying objects in images or videos sequences [4,7]. Humans perform this task effortlessly and instantaneously; however, it remains an exceptionally difficult activity to be outlined from the machines. The object recognition problem is closely engaged to the segmentation problem, while it can also be defined as a labelling problem based on model of known objects. Figure 1 illustrates an example of object recognition for two images.

Image classification is the task of categorizing images into classes according to their depicted contents [13,5]. Large-scale image classification has received serious interest from the computer vision and machine learning communities [1,3]. Figure 2 demonstrates a conventional example of classifying two images.

If we conduct a further more detailed inspection of the Figures 1-2, the results concerning the two typical aforementioned vision-based problems can easily be deduced for both set of images. However if we were about to answer the following question: *is there a human right being violated in these images?* will both set of images draw the same conclusion? Apparently the answer to this question is



Fig. 2. Example of image classification results for two images. They are both carrying the label of 'camp', but it is clear that the image on the right cannot be considered for the task of recognising human rights violations.

totally different for each image. Regarding Figure 1, while both images contain a gun, only in the left image there is a human right being violated which is 'right to life' and 'freedom from torture and inhuman treatment'. Figure 2 demonstrates classification results with both images being labelled as 'camp'. However, the actual situation is quite contrasting with the right image displaying a summer camp, while on the other hand the left image depicts a refugee camp.

For these reasons, detecting and identifying human rights violations through digital images cannot be considered as a trivial problem, since individual tasks need to be combined in order for the system to be able to accurately answer the question whether there is a human right being violated in a specific image or not. Previously this role was played by qualified human rights imagery analysts. Recently, representation learning methods and essentially convolutional neural networks (CNNs) [6] have enjoyed great success in context-based image understanding. Consequently, obtaining effective high-level representations from large-scale web images has become increasingly important, while a key question arises in the context of human rights understanding: how will we gather this structured visual knowledge? This section describes the difficulties encountered when collecting those high-level representations for the task of detecting human rights violations.

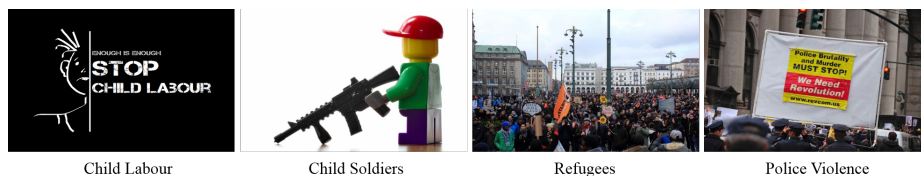


Fig. 3. Examples of irrelevant images returned from Flickr photo-sharing website alongside the given query term.

3.2 Preliminary Research and Discussion

To our knowledge, there is no reference point in standardized dataset of images and annotations regarding human rights violations. The basic component of our intended research is to construct a well-sampled image database in the domain of human rights understanding before exploring the capacity of CNNs at detecting and identifying human rights violation through digital images. The first serious issue arise when Flickr photo-sharing website was chosen in order to collect images related to specific keywords assembled by experts on the field of human rights. Figure 3 displays an example of irrelevant results returned for some given queries. There have been cases where the given keyword was ‘armed conflict’ and most of the returned images had to do with military parades. Another similar example was when the given keyword was ‘refugees’ and the returned results included protesting campaigns for refugees, something that may be consider close to the keyword but can not serve our purpose. Google and Bing search engines were chosen in order to overcome the aforementioned difficulties. However, quickly it became apparent that there was a huge number of irrelevant results returned for the given queries especially when compared with typical inquires such as ‘car’, ‘aeroplane’, ‘bike’, ‘dog’ etc. Table 1 illustrates the percentage of relevant images from the total number of images retrieved for each query concerning four human rights violations categories compared to everyday image queries. Images were downloaded for each class using a python interface to the Google and Bing application programming interfaces (APIs), with the maximum number of images permitted by their respective API for each query term. While the process may be repeated several times in order to get additional images, for the purpose of this comparison a single attempt is presented. This is due to the limited amount of images returned for the rest of the attempts that were conducted with regard to human rights violations. It is evident from Table 1 that in order to obtain real-world images, which can be regarded as appropriate for the creation of a dataset in the human rights violations detection context, greater effort and vast time are required compared to other conventional datasets.

4 Conclusions and Future Work

Recognising and identifying human rights violations through digital images is a unique and challenging problem in the field of computer vision. The overall goal of this paper is to investigate by what means semantically meaningful content must be extracted from digital images in order to highlight those that can be considered for human rights monitoring purposes. The results of this preliminary research indicate that obtaining suitable high-level representations in the context of human rights violations, is a demanding task. Additionally, what is clear from investigating the capacity of real-world images in the context of human rights understanding is that human rights advocates can be supported in the difficult tasks of verifying and tracking human rights violations, eye-witness testimony, prosecuting and other statements on human rights violations in greater extent

Table 1. Statistics of the image collection procedure from search engines. First column displays the query terms. Next, the number of retrieved images alongside the number of applicable images is presented. The last column indicates the quantitative relation between those two sets of images.

Query Term	Retrieved Images		Relevant Images		Ratio	
	Google	Bing	Google	Bing	Google	Bing
child labour	99	137	18	5	18%	3.64%
child soldiers	176	159	31	13	17.61%	8.17%
police violence	149	232	10	16	6.71%	6.89%
refugees	111	140	10	39	9%	27.85%
aeroplane	170	137	150	135	88.23%	98.54%
car	145	128	123	124	84.82%	96.87%
dog	105	132	101	129	96.19%	97.72%

than previous high cost approaches. Inspired by the great success of representation learning methods and principally CNNs, our future research will focus on tackling the untrodden problem of detecting and recognising human rights violations throughout real-world images by exploiting these novel techniques.

Acknowledgements

This work was supported by the Economic and Social Research Council [grant number ES/M010236/1].

References

1. Bengio, S., Weston, J., and Grangier, D. (2010). Label embedding trees for large multi-class tasks. In *Advances in Neural Information Processing Systems*, pages 163–171.
2. Chen, F. and Neill, D. B. (2015). Human rights event detection from heterogeneous social media graphs. *Big Data*, 3(1):34–40.
3. Deng, J., Berg, A. C., Li, K., and Fei-Fei, L. (2010). What does classifying more than 10,000 image categories tell us? In *European conference on computer vision*, pages 71–84. Springer.
4. Fergus, R., Perona, P., and Zisserman, A. (2003). Object class recognition by unsupervised scale-invariant learning. In *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, volume 2, pages II–264. IEEE.
5. Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105.
6. LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., and Jackel, L. D. (1989). Backpropagation applied to handwritten zip code recognition. *Neural computation*, 1(4):541–551.

VIII

7. Lowe, D. G. (1999). 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.
8. Marx, A. and Goward, S. (2013). Remote sensing in human rights and international humanitarian law monitoring: concepts and methods. *Geographical Review*, 103(1):100–111.
9. Marx, A. J. (2013). A new approach to detecting mass human rights violations using satellite imagery. *United States Holocaust Memorial Museum*.
10. Matheson, K. (2016). *Video as Evidence Field Guide*. witness.org, Brooklyn, NY 11217, 1 edition.
11. Notley, T. and Hankey, S. (2012). Human rights defenders and the right to digital privacy and security. *Human Rights and Information Technologies: Trends and Consequences of Use*, pages 157–175.
12. Office of the High Commissioner for Human Rights, O. (2001). Manual on human rights monitoring.
13. Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.