Lyndon J B Nixon
MODUL University Vienna

An Online Image Annotation Service for Destination Image Measurement

This research note reports on the first, to the best of the author's knowledge, release of an online image annotation service for destination image measurement. Destination Marketing Organisations (DMOs) today, while actively data mining textual content for insights into visitor sentiment towards their destination or the most popular topics or themes of visitors at that destination, increasingly face usage of digital imagery or videos - yet non-textual content is not as easily 'understood' by machines to provide the same insights. The recent emergence of online services for image annotation might be of value to DMOs and researchers but their genericity means that to date e-tourism researchers continue to use manual approaches to media annotation, unable to scale to larger data sets and inconsistent across efforts with respect to chosen visual categories and concepts. We present here initial results which indicate that researchers and organisations could use an online service tuned specifically to the detection of visual concepts related to destination image, allowing them to annotate media at greater scale and analyse and compare results according to a common annotation vocabulary, helping us progress further in this exciting new area of e-tourism research and Tourism Intelligence.

Key words: media mining, media analysis, media annotation, visual analysis, concept

detection, image annotation, destination image, tourism intelligence.

Lyndon Nixon
Department of New Media Technology
MODUL University Vienna
Am Kahlenburg 1
1190 Vienna
Austria
Email: lyndon.nixon@modul.ac.at

Lyndon Nixon is Assistant Professor in the Department of New Media Technology at the MODUL University Vienna. His research interests cover the analysis and description of online media and the use of this description in the organisation and presentation of media assets, with a focus on enhancing media value for tourism organisations ("Visual Destination Image"), newsrooms and journalists (video verification), and for broadcasters (Linked Television, TV Intelligence).

Introduction

Digital travellers are more likely today to learn about or check out potential destinations through photos on Instagram, videos on YouTube or visual pins on Pinterest than to read text-based travel blogger entries, travel guides or DMO websites. The shift in digital consumer behavior towards (audio)visual content on the Web raises a new challenge for tourism stakeholders who traditionally have performed data analysis for market research and prediction based on textual and statistical data (Tourism Intelligence). Insights into how destinations and tourism offers are being presented to online consumers will only capture the whole story if the (audio)visual content can be analysed and understood in the same way as today's tourism intelligence solutions can perform with text. Modern advances in computational understanding have enabled significant progress in computer systems that can accurately identify concepts in visual content and label frames according to emotional characteristics, objects and events. Such powerful visual annotation capabilities are even made available publicly via Web services, meaning that functionality that has long been only accessible to very few based on highly complex and expensive computer systems is now a possibility for any business who identifies a business need for it. DMOs and other tourism stakeholders could benefit from the use of state of the art media annotation in order to introduce or extend their text-based Tourism Intelligence capabilities. However, existing "out of the box" solutions are too generically focused on the objective identification of what objects are visible in an image or video and e-tourism research has to date continued to rely on manual annotation of small media datasets, an approach which can't scale to provide tourism intelligence in organizations. In this research note, we present a first implementation of a visual concept classifier tuned to the measurement of destination image, and discuss its potential uses in future e-tourism research as well as contribution to a new generation of Tourism Intelligence solutions in our digital, multimedia world.

Literature Review

Tourism marketers have long been interested in the “destination image” that their audience has of the destination and how to influence that image through their own marketing content. The analysis of non-textual content for tourism marketing began with tourist photographs taken at a destination. With the Web and social media providing free public and global distribution channels for content, combined with the ease of creation of digital image and video assets, tourism media about destinations is now being created at a huge scale, by a very large number of smaller channels of travel blogs and individual travellers. Since travellers now also increasingly use social networks as a source of information about destinations [Xiang & Gretzel, 2010], more recent studies have turned to destination image from online media. [Stepchenkova & Zhan, 2013] compared DMO and Flickr photos along 20 destination attributes, constructing maps of the projected and perceived images of Peru. An aggregated destination image can be formed following the procedure in [Stepchenkova & Li, 2012, Stepchenkova & Li, 2014] by calculating the frequency of occurrence of destination image attributes in the sample. [Fatanti & Suyadnya, 2015] looked at how Instagram creates a tourism destination brand, analysing the promotional value of Instagram through "photo elicitation interview (PEI)". Tourism research has looked at the use of Instagram in destination marketing (e.g. [Hanan & Putit, 2014]). [Nixon, 2017] tested if Instagram content can positively influence a person's perception of a destination, referring to the destination image formed or influenced by visual content as the “visual destination image”. As the number of scientific publications about “visual destination image” in the e-tourism domain has grown, [Picazo & Moreno-Gil, 2017] surveyed the body of literature in the field and noted importantly the inconsistency in annotation methods and models, preventing scientific comparisons between works or re-use of results. [Nixon, 2018] raised already the question if online image annotation services could be used to support “visual destination image” measurement, noting the lack of transparency about the known concepts in their 'black boxes' and the need for additional processing of the annotation results, concluding that e-tourism research needs its own annotation service tuned to the requirements of destination image measurement.

Theoretical Model of Visual Destination Image

The image annotation service needs to be built upon a clearly defined theoretical model for visual destination image measurement. It has already been noted that a shortcoming of e-tourism research to date in this field has been the lack of a common approach or model. Beginning with theories of destination image, we follow the original concept presented by [Beerli, 2004], where according to her model of the formation of destination image, (1) the visual destination image is clearly a 'cognitive' image as it is based on what is seen, rather than 'affective' based on what is felt, and (2) our visual destination image measurement is based on the subject's consumption of visual imagery from various secondary information sources (with research to date focused on the Web, social networks or photography) rather than the direct or 'primary' experience of being there. Therefore we acknowledge that visual destination image is a subset of the subject's overall destination image.

We therefore define Visual Destination Image as a cognitive model based on visual concepts that can be objectively recognized, i.e. the theoretical equivalent of asking someone to close their eyes and imagine being in a certain destination, then asking them what they "saw". Such an image can not be formed from any single image (or video). In fact, a complete model would need to be built up from all visual inputs regarding the destination to the cognitive model in an individual's mind, whether remembered consciously or not. In our research, as in other e-tourism works, the measurement of the visual destination image is restricted to those visual inputs from a selected source, e.g. tourist photography such as on Flickr, a social network feed such as Instagram, or a Website such as that of the DMO. An open research question is how many images need to be considered in order to come to a visual destination image representative of how the destination is really seen, with research to date using comparatively small datasets due to the need for manual annotation of every image.

Likewise, while visual destination image could be calculated for one individual (i.e. taking only the images that they see), research tends to look at the destination image as a shared construct for an entire audience, e.g. as represented by a DMO's online channel or by recent UGC. This may be valid for measuring how the DMO (or other stakeholder) is communicating the destination image visually, but overlooks how the consumer may interpret the visual imagery differently, e.g. different people are likely to focus on different aspects within images. On the other hand, trying to represent the visual destination image for each and every individual would leave us with a model too complex to understand and re-use, therefore we suggest as best practice the segmentation of the audience of visual destination imagery in order to at least reflect that different audience types will build up diverging visual destination images. In order to evaluate the use of an image annotation service which supports our concept of Visual Destination Image, we will consider two hypotheses:

H1. An image annotation service specifically trained on Visual Destination Image can accurately annotate cognitive destination image in visual media better than 'out of the box' generic online services (baseline); and

H2. The annotation of cognitive destination image in visual media is useful for defining the Visual Destination Image in e-tourism research.

Motivation and Method

While tourism research has only partially covered the insights that images or videos could provide and generally examined small numbers of media assets which have been manually annotated in advance by human experts, advances in computer vision (what concepts computers can "see" in an image) offer tourism organisations the possibility to accurately annotate their own images as well as images being shared by visitors about their destinations, in order to gain deeper and valuable insights into the common topics and interests of visitors with respect to their destinations, as well as adapt their own (media) marketing campaigns appropriately.

First of all, a common, shared annotation vocabulary is necessary. In [Nixon, 2017], a subset of [Beerli, 2004]'s factors influencing destination image were used in order to define a set of cognitive attributes for the annotation of destination image in visual media. For each of the nine dimensions of [Beerli, 2004] we extracted the attributes best suited to cognitive image (tangible enough to be captured in an image), leading to an initial list of 53 classes from 7 of the dimensions. In some cases, we were more specific than the Beerli attributes (e.g. from "Flora and Fauna" we extracted the 3 classes 'Plants & Flowers', 'Animals' and 'Trees') since we want to distinguish between different attractions at the destinations (e.g. a botanical garden, a zoo or a jungle). Three further classes for visual objects were added during an initial manual annotation, identifying relevant visual concepts in the images not in our original list: modern buildings, cycling and boating. So our vocabulary for visual destination image is currently 56 classes in 7 dimensions.

In a previous work ([Nixon, 2018]) we analysed 'out of the box' image annotation services available online: CERTH's Image Concept Detection Service (multimedia.iti.gr/mediamixer_images/demonstrator.html [last checked Sept. 20, 2019]) and

IBM's Watson Visual Recognition Service (ibm.com/watson/services/visual-recognition/demo/#demo [last checked Sept. 20, 2019]). We found that, in the dimensions of most significance to the destination, the results of the online services varied the most. Evaluation results were lower than the 'state of the art' figures reported for these services in scientific literature, since here evaluation is made against image collections from previously known domains and they are pre-trained for the concepts they expect to encounter. Therefore, we decided to train our own visual classifier, using the destination image vocabulary described above.

There are different implementations available for building a visual classifier based on AI technologies, extending computer vision algorithms such as OpenCV with machine learning or neural networks. Generally speaking, the computer vision algorithms provide image processing techniques such as colour and shape (edge) feature extraction. In machine learning, sets of images with their 'correct' classifications are fed to a training model (supervised learning approach) and the system learns to correlate the image features (like colours and shapes) extracted by the computer vision algorithm with the provided classifications, such that it can now correctly classify new, previously unseen (and unclassified) images. More complex training models through neural networks combined with very large training data sets have led to the emergence in recent years of impressively accurate, generically applicable image annotation services, provided by the different large Internet companies (Google, Microsoft, IBM etc.). Open source or smaller commercial implementations still require training with the right data, but can allow users to operate independently of an API controlled by one large Internet company. Given the lack of the same scale of data available in such cases, such an independent visual classifier is more effective if focused on a specific domain.

We decided to prototype our idea for a visual destination image classifier using MachineBox.io, a company that provides out-of-the-box machine learning components (called 'boxes') that can be deployed and trained locally. These boxes are available free for non-commercial use. MachineBox.io provides an initial 'default trained' visual classifier known as TagBox and tools to train that classifier with your own annotated images. The free option is limited to 100 visual 'tags' which is more than we need for our destination image classification. We found TagBox came pre-trained with ImageNet concepts, which can be seen as a superset of our set of visual classes, and outputs for an image a list of concepts with a confidence score. However, this vocabulary is similarly generic and can not replace training with our destination image classes; on the other hand, we find some ImageNet classes can be deemed equivalent to our own. To provide a threshold for 'positive' visual classification, we use a threshold of 50% confidence to consider a concept as having been detected in an input image. For training, we selected 37 visual concepts in our vocabulary for which we found training images in ImageNet, downloaded the URL lists for each selected concept (called 'synset'), and used these to train the classifier, using 20 images from each list. In doing so, images were filtered to those we considered relevant to Destination Image, e.g. for "water" bodies of water were used but not close-ups of drops of water. We withheld from each training set four images which were used for testing, i.e. after training, we validated the concept detection by checking that the respective concept would be detected automatically in the four test images (detection being positive if the concept appears in the Tags output by TagBox with a confidence score above 50%).

Evaluation Results

For our first hypothesis, we compare the annotations of different online services – the image annotation service introduced in this paper and a generic 'out of the box' service - with our ground truth annotation. We re-use the IBM Watson service already used in [Nixon, 2018], as it provides an easy to use Web interface demo to annotate any image and can be considered state of the art AI technology. To measure the accuracy of the annotations, we will use precision and recall calculations for classification tasks, where our task is to classify the image in terms of the (destination image-related) visual concepts it presents to a viewer, as well as f-measure as the weighted harmonic mean of both precision and recall, thus providing a combined measure of the annotation service's accuracy.

For this experiment, we collect again the most recent 25 photos posted by the Vienna Tourist Board at instagram.com/viennatouristboard (as of Sept. 20, 2019). We manually annotate them according to our destination image annotation vocabulary and then also use each service to generate image annotations, taking as valid all concepts returned with a confidence level > 50% and directly aligned to our destination image vocabulary. The results are shown in Table 1. While for this sample Watson performed marginally better, we must say it is impressive to get such a similar result with our service which was trained on 16 images for each destination image class, surely a far smaller training dataset compared to what IBM Watson will have used. This indicates potential for our service but the need for further training. We note that we have used ImageNet images which are also used in other visual classification training, therefore we prepare our service to perform the same on visually similar images. For a next phase, we will train the service further on pre-annotated images from DMOs, which will provide a visual training much closer to how tourism organisations represent destinations in terms of destination image (e.g. as previously mentioned, ImageNet provides a wide range of images for the class 'Water' such as water drops whereas we train exclusively on clearly represented bodies of water such as lakes, rivers or seas which are representative of 'Water' in a visual destination image).

Table 1. Comparative accuracy of online image annotation services

Tool	*Precision*	*Recall*	*F-measure*
CERTH	*0.75*	*0.48*	***0.59***
Watson	*0.67*	*0.57*	***0.62***
Our image annotation service	0.67	0.5	**0.57**

Table 2. Sample of image annotation results

Image URL	*Ground truth annotation*	*Our service*	*IBM Watson*
https://www.instagram.com/p/B2jOrTBoidZ/	Historical building	-	Landscape, **historical building**
https://www.instagram.com/p/B2gobR6oZgb/	Religion, historical building	**Religion, historical building**	**Religion, historical building**
https://www.instagram.com/p/B2eDnO7IePP/	Historical building	Religion, **historical building**	**Historical building**
https://www.instagram.com/p/B2be7uVoCB8/	Trees	Monument, **trees**, art	-
https://www.instagram.com/p/B2Y9t33oNlV/	Religion, historical building, water	**Historical building**	**Religion, historical building**

Table 2 shows a sample of 5 of the images, and the differing destination image annotations returned by the services alongside our manual 'expert' annotation. In bold are the 'true positives' – the correctly annotated destination image classes by automatic services, compared to the ground truth expert annotation.

For the second hypothesis, we consider the destination image measurement from the visual annotation. Following other works such as [Stepchenkova & Li, 2012, Stepchenkova & Li, 2014], we assume that the comparative frequency of occurrence of the categories in the media annotations may act as determinants for the visual destination image. To simplify the model, we aggregate the visual class occurrences into the 7 top level categories used by Beerli. As a result, any visual destination image may be represented as a vector made up of frequency measurements for the 7 destination image categories. This allows us a means to mathematically calculate distances between different destination images, and can be visualized by various means such as histograms.

Figure 1 shows the visual destination image derived from the ground truth and from the annotations by our image annotation service and IBM Watson. The ground truth indicates that destination images posted by the Vienna Tourist Board promote most strongly Vienna's natural resources and its cultural/historical offer (interestingly the same result from 2017, so Vienna's DMO is very consistent in the visual destination image it promotes!). Also, as last time, the destination images derived from the automatic services underrepresent natural resources in imagery, so clearly this is still an area to work on in destination image classification. Comparing the destination images below, it can be seen the overall image from our service is closer than IBM Watson's to the ground truth in that it also reflects the dominance of the two classes 'Natural Resources' and 'Culture, History and Art', whereas Watson overrepresents General Information (orange) and Leisure and Recreation (yellow).

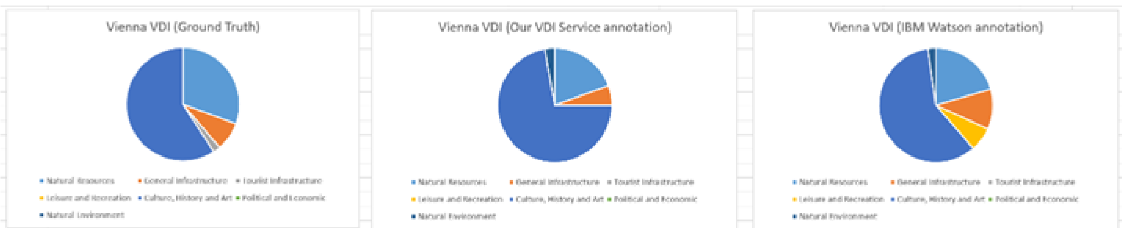

Figure 1. Comparative measurement of Visual Destination Image across tools

As a result, we can circumstantially say that both hypotheses could be accepted on the basis of further training and refinement of our approach. Our image annotation service has been demonstrated to accurately (- as accurate as state of the art systems -) annotate images according to typical destination image characteristics and those annotations may be used to model a 'visual destination image' of a destination that can then be compared with other results.

Conclusions

On the basis of this study, based on a small sample of Instagram photos posted by the Vienna Tourism, we may draw some initial conclusions about the contribution of an online image annotation services for visual destination image. We focused on the measurement of destination image, a common model for tourism stakeholders to consider how a destination is being presented. While text analysis tools have matured and are being increasingly used by DMOs for this task, multimedia analysis is a "brave new world". We indicated that off-the-shelf solutions are not yet performing as well in the tourism domain as they do in their evaluations reported in the research community, where they are pre-trained on image collections from previously known domains. Contemporary e-tourism research in destination image as applied to photographs and other imagery has not yet been able to benefit from the developments in AI and computer vision that suggest that multimedia analysis could become as available to tourism stakeholders as text analysis has been since decades.

To remedy this situation, we trained an online visual classifier with destination image concepts. We benchmarked our destination image-specific image annotation service against a generic, leading state-of-the-art service to show that we provide accurate image annotations with respect to visual destination image. We also considered if the resulting visual destination image model formed from annotation of a larger set of images could be useful in e-tourism research, where questions are still being asked such as how destinations are being represented by visual content, to what extent this visual representation matches DMO's own communication strategies, or influences consumer behaviour (e.g. intention to visit), or compares to other destinations, or varies across audiences. A cursory test suggested that our annotation service can be valuable in answering such research questions but this remains to be further proven by its use in providing researchers with the accurate and appropriate image annotations they need for their research. To train the service, we would need more labelled destination image photography, which could be done by crowdsourcing annotations at scale and using cross-annotator agreement to resolve the typical challenge of annotators having different interpretations of the visual imagery.

As both providers and consumers of destination information use more image and video content, useful services for touristic multimedia annotation will be vital for accurate tourism intelligence in the future. We hope this first experiment with training a classifier to annotate images according to destination image can help drive more research into visual destination image, its measurement and use in e-tourism. As a Web service providing classification trained to the domain of Visual Destination Image, this could potentially support larger scale evaluations of destination marketing and traveller perceptions than the current e-tourism research has been able to perform.

References

Beerli, A. & Martín, J.D. (2004). Factors Influencing Destination Image. Annals of Tourism Research, 31(3), pp. 657-681.

Fatanti, M.N. & Suyadnya, I.W., (2015). Beyond User Gaze: How Instagram Creates Tourism Destination Brand? Procedia – Social and Behavioral Sciences, 211, pp. 1089-1095.

Hanan, H., Putit, N., (2014). Express marketing of tourism destinations using Instagram in social media networking. In Hospitality and Tourism: Synergizing creativity and innovation in research, pp. 471-474.

Nixon, L., Popova, A. and Önder, I. (2017). "How Instagram influences Visual Destination Image: a case study of Jordan and Costa Rica", ENTER2017 eTourism conference, Rome, Italy

Nixon, L. (2018). "Assessing the usefulness of online image annotation services for destination image measurement", ENTER2018 eTourism conference, Jönköping, Sweden, January 2018.

Picazo, P., Moreno-Gil, S. (2017). "Analysis of the projected image of tourism destinations on photographs: A literature review to prepare for the future", Journal of Vacation Marketing, pp. 1-22, November 2017.

Stepchenkova, S., Li, X., (2012). Chinese outbound tourists' destination image of America: part II. Journal of Travel Research, 6, pp. 687-703.

Stepchenkova, S., Zhan, F. (2013). Visual destination images of Peru: Comparative content analysis of DMO and user-generated photography. Tourism Management. 36. pp. 590-601.

Stepchenkova, S., Li, X., (2014). Destination image: Do top-of-mind associations say it all? Annals of Tourism Research, 45, pp. 46-62

Xiang, Z., Gretzel, U., (2010). Role of social media in online travel information search. Tourism Management, 31, pp.179-188