Jia, S., Lansdall-Welfare, T., & Cristianini, N. (2016). Time Series Analysis of Garment Distributions Via Street Webcam. In A. Campilho, & F. Karray (Eds.), Image Analysis and Recognition: 13th International Conference, ICIAR 2016, in Memory of Mohamed Kamel, Póvoa de Varzim, Portugal, July 13-15, 2016, Proceedings. (pp. 765-773). (Lecture Notes in Computer Science; Vol. 9730). Springer. DOI: 10.1007/978-3-319-41501-7_85

Peer reviewed version

Time Series Analysis of Garment Distributions Via Street Webcam

Sen Jia, Thomas Lansdall-Welfare, and Nello Cristianini

Intelligent Systems Laboratory, University of Bristol, Bristol, UK,
firstname.lastname@bris.ac.uk

Abstract. The discovery of patterns and events in the physical world by analysis of multiple streams of sensor data can provide benefit to society in more than just surveillance applications by focusing on automated means for social scientists, anthropologists and marketing experts to detect macroscopic trends and changes in the general population. This goal complements analogous efforts in documenting trends in the digital world, such as those in social media monitoring. In this paper we show how the contents of a street webcam, processed with state-of-the-art deep networks, can provide information about patterns in clothing and their relation to weather information. In particular, we analyze a large time series of street webcam images, using a deep network trained for garment detection, and demonstrate how the garment distribution over time significantly correlates to weather and temporal patterns. Finally, we additionally provide a new and improved labelled dataset of garments for training and benchmarking purposes, reporting 58.19% overall accuracy on the ACS test set, the best performance yet obtained.

1 Introduction

The idea of observing collective behaviour on a large scale goes back to at least the 1930s in the social sciences [12], where the general goal was to record and study social habits, the spread of fashions, and even the (anonymous) opinions of the masses. Indeed, the study of how certain habits and behaviours arise and spread has continued to be of great interest to marketing and social sciences alike.

While classical tools were not much different than those of anthropology, involving interviews, observations and questionnaires, it has recently become possible to access the aspects of collective behaviour that are online. For example, the analysis of Twitter content has been used to detect large-scale mood shifts [9], as well as detecting phenomena such as flu outbreaks [8]. Major shifts in culture have also been detected in book content [11] and in musical styles [10] by data-driven approaches. However, observing collective behaviour in the physical world requires a different approach, one which can take advantage of the widespread distribution of cameras and other physical sensors within our society. While anonymous observation of collective behaviour certainly has implications for

ACS Category	# Images	Fashionista Category	# Images	Merged Training Set	Total
blouses	896	blouse	13,808	blouse	14,704
cloak	7,061	-	0	cloak	7,061
coat	9,061	coat	4,738	coat	13,799
jacket	9,366	hoodie, jacket, cardigan	23,879	jacket	33,245
long dress	10,090	-	0	long dress	10,090
polo shirt, sport shirt	780	-	0	sports shirt	780
robe	5,799	-	0	robe	5,799
shirt	1,425	-	0	shirt	1,425
short dress	4,285	romper, dress	36,035	short dress	40,320
suit, suit of clothes	6,054	blazer	12,651	suit	18,705
sweater	5,209	sweater	8,934	sweater	14,143
jersey, T-shirt, tee shirt	1,426	t-shirt, shirt	28,950	T-shirt	30,376
undergarment, upper body	5,538	bra	757	undergarment	6,295
uniform	3,353	-	0	uniform	3,353
vest, waistcoat	750	vest	6,959	vest	7,709
Total training images	71,093		136,711		207,804

Table 1. A training set created by merging garment categories from the ACS and Fashionista datasets is used to train a garment classifier.

mass surveillance, it can also shed light on interesting trends that might be found in various aspects of society.

Previous work has used this idea to monitor activity in smart cities [1], using sensor meta-data and semantic annotations of street camera and microphones to detect changes. A further study [13] used webcams to detect events in New York, such as a parade, by fusing information from visual (CCTV) and social (Twitter) sensors in the same location. In those approaches the emphasis is on the combination and modelling of multiples streams of sensor data, including cameras and social media content, to detect any interesting patterns or violations thereof.

In this paper, we show how recent advances in deep learning applied to computer vision, combined with webcam data freely available on the web, can be used in a similar way to reveal patterns in what people are wearing, and how this relates to the temporal and weather patterns in the same location. More specifically, we train a Convolutional Neural Network (CNN) [7] on a new, combined and annotated dataset of clothing images generated from the "Apparel Classification with Style" (ACS) [2] and "Fashionista" [14] datasets, achieving a new record of 58.19% on the ACS test set. We used this classifier to find garments items in 243,470 images taken on the street in New York over 420 days before analyzing patterns of pedestrian flow and choice of clothing relative to the time and the local weather.

Our approach further exemplifies the opportunities for a new sort of mass observation, one which can discover large-scale patterns in society by combining multiple sources of information, including webcams, social media, sensor data, and more. This can (and should) be done without the need to use any personal data, if automated anonymisation techniques are used and the data is aggregated, as performed in this study.

This paper makes the following contributions:

1. We compiled a new training set for garment classification that we show achieves state-of-the-art performance when combined with deep learning networks,
2. We analyze real-world images from a street webcam in Williamsburg, New York, showing that there are trends in the pedestrain flow and what people are wearing based upon the time of year,
3. We show that the choice of garments found in the images can be partially explained by the local weather.

2 Methods and Data

2.1 Garment Training Data

In order to train a classifier that can recognise instances of different garments within images, we first created a training set by merging together two of the most popular, publicly available datasets for garment classification, namely the ACS [2] and Fashionista [14] datasets. Fashionista contains 158,235 images collected from the "Chictopia.com" fashion website, organized into 53 garment categories by users when they upload images, while the ACS dataset contains 88,951 garments images organized into 15 garment categories crawled from the web based upon ImageNet [3]. The ACS dataset is provided split into a training set containing 71,093 images, which we use here, and a separate test set containing 17,858 images which we hold out to use to validate our garment classifier in Section 2.2.

Using the 15 garment categories provided by the ACS dataset, we manually selected the closest corresponding categories from the Fashionista dataset, as shown in Table 1, allowing us to merge the data into a single training set. To ensure that there was no intersection between our newly created merged training set and the ACS test data set, we performed the following test. We pre-processed each training image, following the procedure in [2], by resizing the largest side to 320 pixels and normalising the image histogram. Each training image was then compared to each test image using a pixel comparison, finding that there was no intersection between the two datasets.

2.2 Training a Convolutional Neural Network

Typically, the identification of garments within images requires accurate pose estimation in order to extract what a person is wearing [15]. However, this approach is often prohibitively expensive for real-world applications. We take a deep learning approach to garment classification, implementing a convolutional neural network using the AlexNet [7] network in Caffe [6], a popular deep learning library.

Following the procedure of AlexNet [7], we first augmented the training set using multiple-crop and horizontal flips. Using a batch size of 128, we trained the network for 450,000 iterations with a dynamic learning rate starting at 0.01 and scaling 10 times smaller every 100,000 iterations. We take the final snapshot of the classifier after the last iteration as the garment classifier for the experiments in this paper.

We tested the garment classifier on the ACS test set which we held out from the training set, and was found to have zero intersection with the merged training set detailed in Table 1. We found that the overall accuracy of the classifier was a state-of-the-art 58.19% on the ACS test set, a new record for garment classification.

3 Experiments

Once we had a trained garment classifier, we wished to demonstrate on real-world data that the combination of advances in computer vision and freely available webcam data can reveal patterns in real-world trends, such as what people are wearing, and how it relates to the time of day, week and year along with weather patterns in the same location.

3.1 Styleblaster Webcam

Styleblaster is a real-time fashion website that captures images of people walking past a street webcam in Williamsburg, New York during daylight hours (between 7 a.m. and 8 p.m.). The images are uploaded to the website and annotated with a timestamp of when the image was taken. Importantly for this study, the webcam captures an image for every person that walks past the camera without applying any discrimination or filtering, enabling us to study the hourly and daily pedestrian traffic in the area, as well as the fashion trends over time.

For this study, we analyzed 243,470 images from Styleblaster dated between October 2012 and May 2015. We found that occassionally the webcam would capture an empty image (one not containing a person) due to background noise such as roadworks, snowdrifts or changes in illumination. To remove these images from those we analyzed, we applied a person detection algorithm using the fast R-CNN deep network [5] based upon [4]. This process removed 51,299 images, leaving us with 192,171 images containing a person for our analysis.

3.2 Pedestrian Flow Analysis

The webcam images from Styleblaster capture every person it detects walking past on the street without discrimination. This allowed us to analyze the pedestrian flow on the street by counting how many people were detected each hour and on each day, as shown in Fig. 1.

We can see that during weekdays there are two main peaks of activity, one in the morning between 8 a.m. and 10 a.m. and one in the early afternoon between

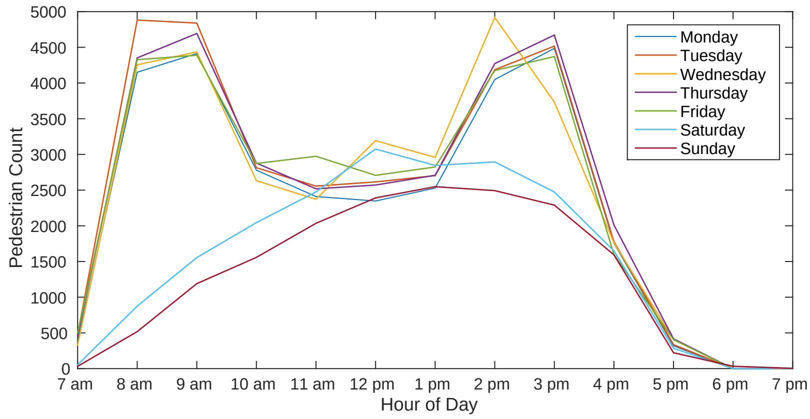

Fig. 1. Total pedestrian count for each hour of the day, separated by day of the week. A clear separation between weekdays and weekends can be seen, with weekdays exhibiting peaks of pedestrian activity between 8 a.m. and 10 a.m., and again between 2 p.m. and 4 p.m.

2 p.m. and 4 p.m. During the weekends, there are no distinguished peaks of activity however, with a gradual increase in pedestrians until just after midday before tailing off again. This suggests that we are detecting people travelling to and from work or school during the weekdays, as the peaks are not found during the weekend.

3.3 Garment Distribution

We further analyzed the distribution of what people are wearing at different times of the year by applying the garment classifier detailed in Section 2.2 to the person bounding box in each of the 192,171 Styleblaster images that contained a person. For each image, we obtained a label and a confidence in the prediction, thresholding the results so that we only consider labels that have a confidence over 0.5, resulting in a total of 36,883 garment detections.

While we report an accuracy across all 15 garment categories, four categories ("blouses", "sports shirt", "shirt" and "vest") received too few positive predictions to meaningfully study the variation in their distribution. We believe that for each of these categories, they received too few predictions either because they have been found to be difficult to classify, both with our classifier and in [2] ("blouses", "vest") or because they are incorrectly classified as another category due to their visual similarity (i.e. "sports shirt" is misclassified as "T-shirt" and "shirt" is missclassified as "jacket").

For the remaining 11 garment categories, we calculated the conditional probability of each month given we have detected a garment, displayed in Fig.2 and Fig. 3 for readability. We can see that there are generally two trends that the garments follow, either peaking in the summer months, for example "short dress",

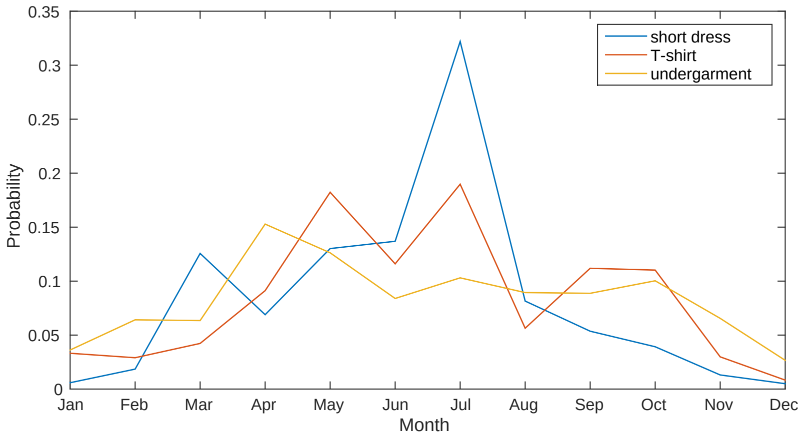

Fig. 2. Conditional probability of each month given that a specific garment was detected.

"T-shirt" and "undergarments"[1] or exhibiting a trough during the summer, as found for most of the rest of the garment types.

3.4 Garment Weather Patterns

We further studied the relationship between the garment distributions and the local weather at that time. We obtained temperature and precipitation data for the closest weather station to Williamsburg, New York from the "National Centers for Environmental Information" website[2] for each day that we had images from Styleblaster, resulting in maximum temperature and precipitation time series similar to those for each garment category. Maxmimum temperature was used due to the images being taken during the daytime, when the maximum temperature for the day is typically reached. We then used multi-dimensional scaling (MDS) to embed the data in a space based upon the similarity of the time series' while preserving distances as much as possible, resulting in a dimensionality reduction of the data.

In Fig. 4 we visualize the first two Eigenvectors from MDS, with the geometric distance indicating the correlation between the different garment categories and the weather information. We found that "T-shirt", "undergarment" and "short dress" are all close to the temperature, indicating that they correlate well with the maximum daily temperature, while the other garment categories are more distant, with "coat" and "jacket" being the furthest from temperature. No garment categories were found to be particularly related to the precipitation, with "cloak" being the closest garment category.

[1] We use the label "undergarment" from the ACS dataset for this category, however it more generally refers to more revealing clothing including tank tops, corsets, spandex sportswear etc., along with actual undergarments.

[2] `http://www.ncdc.noaa.gov/`

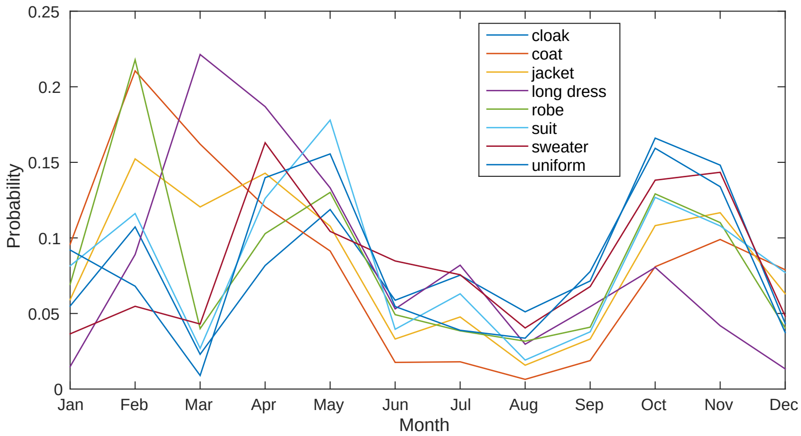

Fig. 3. Conditional probability of each month given that a specific garment was detected.

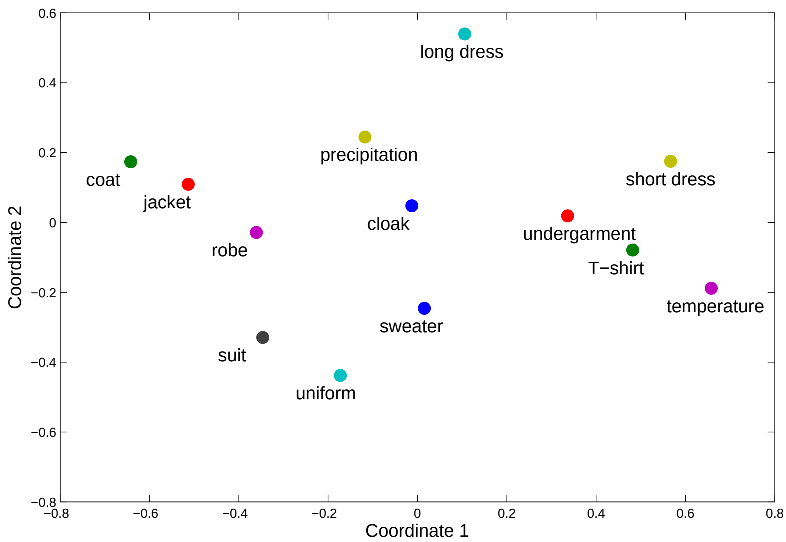

Fig. 4. Visualization showing the similarity of the garment categories with the local daily maximum temperature and precipitation using multi-dimensional scaling.

4 Conclusions

This study shows that computer vision tools based on deep networks can be successfully be applied to street images coming from a webcam, in order to identify garments worn by pedestrians, and generate a time series of garment types. This time series shows significant correlations with weather information and time information.

The possibility of extracting semantic-level information from street cameras opens the possibility of large-scale observation of trends and fashions, which can complement analogous efforts in observation of media content. However it also creates the potential for abuse if it was applied at the individual level. All of our experiments were aimed at observing collective and average behaviour only.

It is easy to imagine a software infrastructure observing hundreds of webcams, as well as other types of web sensors, trying to detect changes, trends, and events. This would surely qualify as an example of Mass Observation to the social scientists of the 1930s, whose goal was stated as the study and recording of the social habits of ordinary people, within an anthropological perspective. The implications of this infrastructure need to be carefully assessed in light of ethical considerations.

References

1. M.-D. Albakour, C. Macdonald, and I. Ounis. Using sensor metadata streams to identify topics of local events in the city. In *SIGIR'15*, pages 711–714, 2015.
2. L. Bossard, M. Dantone, C. Leistner, C. Wengert, T. Quack, and L. Van Gool. Apparel classification with style. In *ACCV'12*, pages 321–335, 2012.
3. J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. ImageNet: A Large-Scale Hierarchical Image Database. In *CVPR'09*, 2009.
4. R. Girshick, J. Donahue, T. Darrell, and J. Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *CVPR'14*, 2014.
5. R. B. Girshick. Fast R-CNN. *CoRR*, abs/1504.08083, 2015.
6. 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.
7. A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *NIPS 25*, pages 1097–1105. 2012.
8. V. Lampos, T. De Bie, and N. Cristianini. Flu detector - tracking epidemics on twitter. In *ECML PKDD*, pages 599–602, 2010.
9. T. Lansdall-Welfare, V. Lampos, and N. Cristianini. Effects of the recession on public mood in the uk. In *WWW'12 Companion*, pages 1221–1226, 2012.
10. M. Mauch, R. M. MacCallum, M. Levy, and A. M. Leroi. The evolution of popular music: Usa 1960–2010. *Royal Society Open Science*, 2(5):150081, 2015.
11. J.-B. Michel, Y. K. Shen, A. P. Aiden, A. Veres, M. K. Gray, J. P. Pickett, D. Hoiberg, D. Clancy, P. Norvig, J. Orwant, et al. Quantitative analysis of culture using millions of digitized books. *Science*, 331(6014):176–182, 2011.
12. J. Moran. Mass-observation, market research, and the birth of the focus group, 1937-1997. *The Journal of British Studies*, 47:827–851, 10 2008.
13. Y. Wang and M. S. Kankanhalli. Tweeting cameras for event detection. In *WWW '15*, pages 1231–1241, 2015.
14. K. Yamaguchi. Parsing clothing in fashion photographs. In *CVPR '12*, pages 3570–3577, 2012.
15. K. Yamaguchi, M. Kiapour, and T. Berg. Paper doll parsing: Retrieving similar styles to parse clothing items. In *ICCV'13*, pages 3519–3526, 2013.