Novelty Detection in Wearable Visual Systems
EXTENDED ABSTRACT

Omid Aghazadeh, Josephine Sullivan and Stefan Carlsson
Compuer Vision and Active Perception laboratory, KTH
{omida,sullivan,stefanc}@kth.se

1. Introduction

Mass production of wearable cameras has made it possible to collect large amount of data each day. Most of this data naturally is of daily repeated activities and this has already started to become the focus of many computer vision scientists. This repeated structure can be useful in learning models for recognition of specific activities, places, objects, etc in supervised [5, 6, 3] or weakly supervised [4] frameworks. Unsupervised learning of this structure has also been explored [1, 2]. Detecting *novelty* - defined as deviation from this repeated structured, the *background*, is the focus of this article.

Our non-parametric approach to novelty detection recognizes the repeated structure which in turn highlights the unique aspects of the data. This potentially has a wide range of applications in 1- lifelogging e.g. automatic memory selection and summarization of events by compressing the background, 2- object recognition e.g. reducing the false positives by filtering out the environment, 3- improving methods which rely on correspondences e.g. dense 3D reconstruction by identifying parts of videos/images which do (not) have correspondences, etc.

Section 2 discusses novelty detection in the temporal domain, giving an overview of [1]. Section 3 discusses novelty detection in the spatial domain, giving an overview of [2]. We will discuss future works and conclude this article in section 4.

2. Novelty Detection in the Temporal Domain

We define novelty within a *(query)* sequence, in the temporal domain, to be the *inability to register* frames of the query sequence to that of the previously stored *(reference)* sequences. Such an approach would require 1- a measure of similarity between a query frame and a reference frame and 2- an aggregation function that estimates the *registrability* of a query frame based on the pairwise similarity measures.

Using such an approach, one could potentially detect a broad range of novelties by defining proper similarity measure and aggregation functions e.g. if the goal is to detect if a place is being visited for the first time (novel place detection): 1- a place similarity measure needs to be defined and 2- the image of the place should not be similar to any other reference image with respect to the place similarity measure. As the proof of concept, we investigate the problem of novel ego-motion detection below.

2.1. Novel Ego-Motion Detection

Here, we exploit the fact that a temporally consistent change in the view point between frames of two sequences suggests similar ego-motions. Therefore, we enforce temporal consistency on the view point changes of two sequences by aligning them, using dynamic time warping, with respect to a view point similarity measure. The view point similarity measure is estimated using Epipolar geometry. Comparing a query sequence and a reference sequence this way, each query frame will be associated with a reference frame - within the reference sequence - and with the corresponding similarity measure.

Using the same approach, we register all reference sequences to the query sequence independently and define the following as the aggregation function: for each query frame, the registerability measure is the maximum similarity measure over all the associated similarity measures to the query frame. This step, associates with each frame in the query sequence, a registerability measure and a smoothed version of this signal[1] is used for novel ego-motion detection. This registerability measure can then be thresholded to detect novelties.

With this approach, we were able to detect novelties (in the ego-motion of the subject wearing the camera) such as running into a friend, visiting a place for the first time or taking an unusual route while travelling from a familiar place to a familiar destination [2].

Data set and Results: We collected 31 sequences of the

[1] Gaussian smoothing is utilized to reduce the noise as the max operator is not smooth

[2] Such an approach would also detect if the environment undergoes a significant change, but as a significant change in the environment is not frequent, we do not focus on that here.

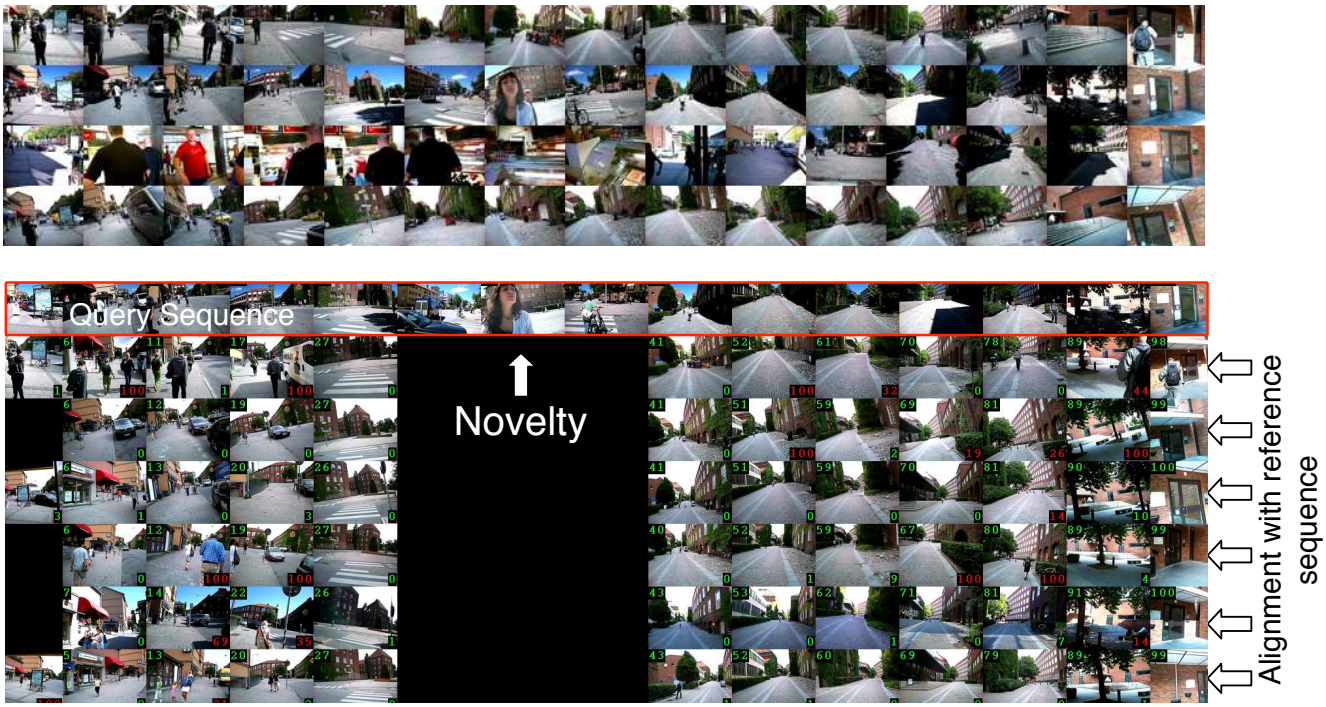

Figure 1. Qualitative results for novel ego-motion detection. **(top):** (highly sub-sampled) raw data before the registration (each row depicts samples from one day). **(bottom):** the subject met a friend and as that resulted in an ego-motion unique in the data set, our system (correctly) detects it as a novelty.

same subject walking from a metro station to work on different days. Sequences are subsampled at 1HZ from the videos that are on average 5 minutes long. Within these sequences, we manually labelled 4 instances of novelty in ego-motion and our method is able to detect those with an Average Precision of 0.96 using 6 reference sequences (see [1] for details). Figure 1 illustrates a qualitative result.

3. Novelty Detection in the Spatial Domain

Here, we consider a similar definition and approach to that of the previous section: we define novelty within a (query) frame, in the spatial domain, to be the inability to register pixels of the query frame to that of the previously stored (reference) images. Consequently, we utilize a similarity measure, an aggregation function and potentially use some constraints to make the approach robust.

The novelties that can be detected using this approach yet again depend on the similarity measure and the aggregation function utilized. As a proof of concept, we investigate the problem of novel object detection below.

3.1. Novel object detection

The problem that we are considering here is to identify pixels within a query frame which belong to the static physical environment. Therefore, the similarity measure - between a pixel in a query frame and a pixel in a reference frame - we consider is a similarity between the physical world points associated with the pixels.

Such a general similarity measure is very hard to estimate, but with the use of additional information e.g. if the two images being compared are of the same physical environment, reasonable approximations to such a similarity measure can be found. Therefore, we only compare images that are approximately of the same physical environment i.e. have approximately similar view points. View point similarity can be approximated by e.g. global geometrical constraints or by using the registration results of the previous section.

Conditioned on the fact that the two images whose pixels are being compared are of approximately the same physical place, we use appearance similarity to define our similarity measure. Similar to the previous section, the appearance similarities are limited to those which agree with some set of correspondences computed via sift flow. The similarity measures are then aggregated via a logistic function learnt in a supervised manner and a smoothness prior is imposed on the solution using a Markov Random Field formulation.

This approach is able to detect novelties in the environment such as people, cars, bicycles, etc in addition to significant changes in the environment e.g. a large poster appearing/disappearing [3].

[3] Significant changes in lighting conditions are likely to disturb any sys-

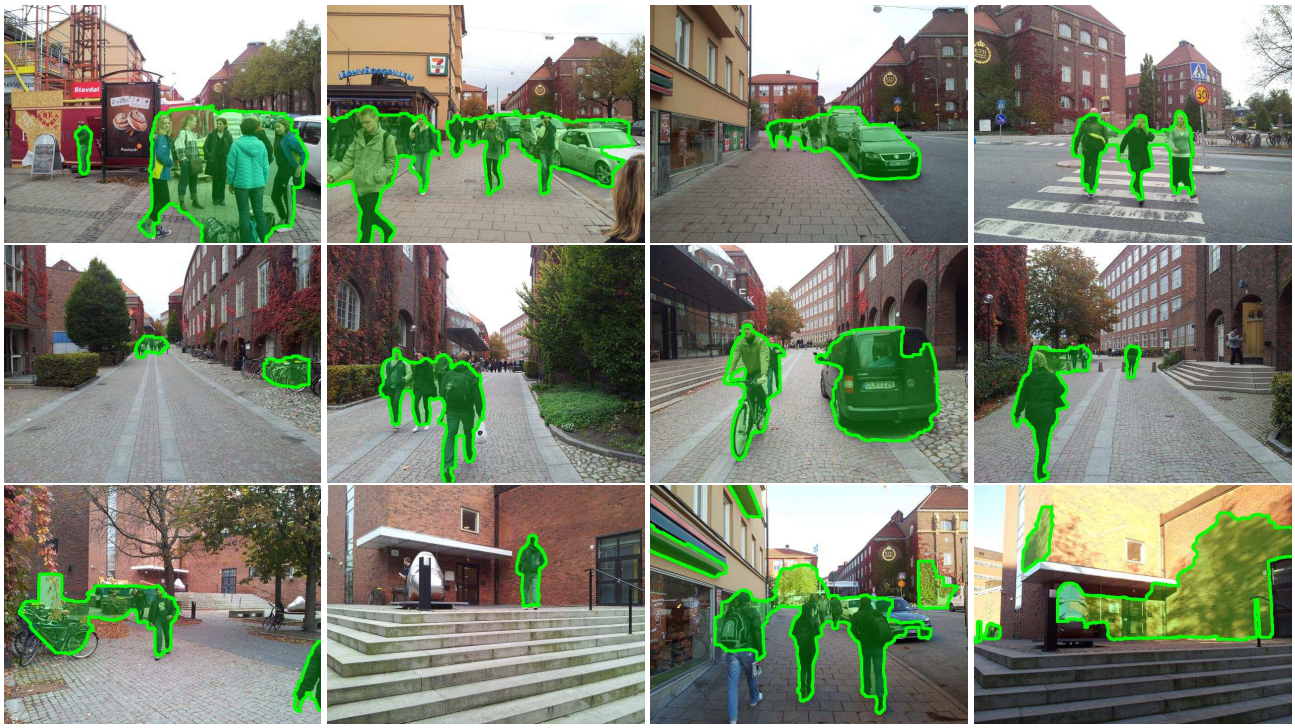

Figure 2. Qualitative results of novel object detection. The segments in green are supposed to not belong to the underlying static physical environment. The last 2 images depict failure cases due to strong changes in illumination conditions and/or textureless novelties occupying textureless background.

Data set and Results: As the wearable camera we used for the previous study had low resolution and distorted images, we used a better camera to collect 12 images of 12 different places (12 x 12 = 144 images total). Each image is manually annotated with novelties rather subjectively.

Quantitative evaluation of our final approach achieves an Average Precision of 0.74 and a pixelwise accuracy of 0.92 (see [2] for more details.). Figure 2 depicts qualitative results of the novel object detection.

4. Conclusions

In this article, we presented a general framework for novelty detection aiming to learn the underlying structure in data sets containing repeated activities - in an unsupervised and non-parametric manner. We showed that novelty detection in our framework comes down to the definition of a similarity measure, an aggregation function over multiple (repeated) instances and some constraints to make the detection robust. We presented novelty detection in two domains: temporal and spatial and investigated a problem for each case as proof of concept.

The crucial fact that allows us to detect novelty is our ability to estimate a structured temporal domain background, by having the subject perform repeated activities day to day. Given more advanced methods for finding similarity and structure, the framework could be extended to more general cases of background and novelty that do not necessarily involve (almost) exact daily repetitions.

Future works include extension of similarity measures/aggregation functions in order to detect other type of novelties and the use of parametric approaches with a reasonable amount of supervision in the novelty detection process.

tem that relies on somewhat reliable local (gradient) structures.

References

[1] O. Aghazadeh, J. Sullivan, and S. Carlsson. Novelty detection from an ego-centric perspective. In *CVPR*, 2011. 1, 2

[2] O. Aghazadeh, J. Sullivan, and S. Carlsson. Multi-view registration for novelty/background separation. In *CVPR*, 2012. 1, 3

[3] A. R. Doherty, N. Caprani, C. O. Conaire, V. Kalnikaité, C. Gurrin, A. F. Smeaton, and N. E. O'Connor. Passively recognising human activities through lifelogging. *CHB*, 2011. 1

[4] A. Fathi, X. Ren, and J. M. Rehg. Learning to recognize objects in egocentric activities. In *CVPR*, 2011. 1

[5] H. Pirsiavash and D. Ramanan. Detecting activities of daily living in first-person camera views. In *CVPR*, 2012. 1

[6] A. Torralba, K. P. Murphy, W. T. Freeman, and M. A. Rubin. Context-based vision system for place and object recognition. In *ICCV*, 2003. 1