Event Detection in Underwater Domain by Exploiting Fish Trajectory Clustering

Simone Palazzo
Department of Electrical, Electronical and Computer Engineering
University of Catania
Catania, Italy
palazzosim@dieei.unict.it

Concetto Spampinato
Department of Electrical, Electronical and Computer Engineering
University of Catania
Catania, Italy
cspampin@dieei.unict.it

Cigdem Beyan
IPAB, School of Informatics
University of Edinburgh
Edinburgh, UK
C.Beyan@sms.ed.ac.uk

ABSTRACT

In this paper we propose a clustering-based approach for the analysis of fish trajectories in real-life unconstrained underwater videos, with the purpose of detecting behavioural events; in such a context, both video quality limitations and the motion properties of the targets make the trajectory analysis task for event detection extremely difficult. Our approach is based on the k-means clustering algorithm and allows to group similar trajectories together, thus providing a simple way to detect the most used paths and the most visited areas, and, by contrast, to identify trajectories which do not fall into any common clusters, therefore representing unusual behaviours. Our results show that the proposed approach is able to separate trajectory patterns and to identify those matching predefined behaviours or which are more likely to be associated to new/anomalous behaviours.

Categories and Subject Descriptors

I.4.8 [**Image Processing and Computer Vision**]: Scene Analysis

General Terms

Algorithms, Experimentation

Keywords

Fish behaviour, trajectory analysis, k-means

1. INTRODUCTION

Although it is one of the most challenging tasks in video analysis, event detection, that is the semantic interpretation of the actions performed by some kind of "actors" in a scene, is the main purpose of many recent applications, such as video surveillance [9, 5], automatic salient scene detection (e.g. in sports events) [8] and the study of animals in their natural habitat [21].

Many approaches have been proposed by the scientific community to address the problem of how to represent events and behaviours and how to develop automatic approaches for identifying complex actions and/or learning "normal" behaviours in order to identify anomalies.

Some approaches (e.g. [11, 3]) identify events by recognising basic motion properties (moving in a certain direction, being still, approaching an object/actor, etc) and composition of such properties into higher-level scenarios ("chasing", "walking, then running", etc). These approaches are able to describe accurately events; however, wrong descriptions might invalidate the results, and anyway such a system would not be able to identify other events than the ones it was trained on.

Other approaches (e.g. [26, 17]) do not try to explicitly map image/motion properties to semantic actions, but apply clustering and machine learning techniques to trajectories in order to learn common motion patterns, which can then be used to identify both usual and unusual events, thus allowing, unlike the previous approach, to possibly discover new behaviours and events. Of course, in this case, it is not possible to use a-priori knowledge on the definition/selection of behaviours.

However, these approaches on event detection are typically designed for people, since of course they represent the most interesting targets for most applications (such as video surveillance). This introduces a great simplification of the general problem, since people's behaviours and motion patterns, besides being easily recognizable and easy to model in terms of low-level actions, are usually "smooth" (i.e. without any sharp changes in speed and direction, whose presence can usually be an indication of unusual activity) and two-dimensional.

A change in the targets – or better, in their motion patterns – to be analyzed has an important effect on the results and even the applicability of the methods typically used for humans. In this work, we tackle the problem of detecting events involving fish in unconstrained underwater footage. Switching the context of the event detection task from people to animals, and even more in the case of fish in a real-life environment, introduces a few major complications:

- Video quality suffers data transmission and bandwidth limitations, due to the difficulties of connecting the underwater cameras to the storage/processing machines.

MAED'12, November 2, 2012, Nara, Japan.

- Trajectory extraction is influenced by both video quality, which severely damages fish appearance and textures, and by the typical characteristics of underwater scenes: sudden light changes due to the gleaming of the Sun on the water, occlusions between fish and/or scene elements, multi-modal background movements which may be mistaken for fish, etc.

- Approaches based on scenario definition are hardly applicable for fish: firstly, their erratic movement makes it difficult to encode a certain behaviour as a sequence of actions; moreover, even by a visual observation of fish movements it is not easy to infer the purpose of their movements. Therefore, using an unsupervised event learning approach is more feasible.

In fish behaviour understanding context, the "events" can be categorized into two fold:

- Fish trajectories not belonging to a set of common paths or not matching a set of common motion patterns, which represent " unusual behaviours".

- Interactions (in terms of similar trajectories at the same time) between fish of the same species (e.g. mating) or different ones (e.g. preying).

In this work we investigated the use of a clustering approach to identify groups of trajectories representing common behaviours, as well as unusual behaviours. The video set we used is randomly selected from the *Fish4Knowledge*[1] project's repository which contains thousands of videos captured from 2010 till now by 10 underwater cameras in several locations of the Taiwanese coral reef, and it contains trajectory of different fish species (9 species) from different time of day in different locations. Our approach uses reliable fish detection and tracking algorithms to extract trajectories; which are given after a pre-processing step as input to k-means [10] algorithm that clusters similar trajectories together. The results show that this approach successfully associates trajectories to visually labeled events (such as being stationary, changing direction, biting from coral) and also identify groups of trajectories which do not match any common behaviour and thus can represent new unseen events (unusual events; to be validated by marine biologists, of course) and provide data for new research. The rest of this paper is organized as follows: Section 2 briefly introduces the current state of the art on event detection in videos; Section 3 explains how fish trajectories are extracted and how clustering is performed; in Section 4 we show the results we obtained in several test scenes; finally, in Section 5 some conclusions and ideas for future developments are drawn.

2. RELATED WORK

Many approaches have been proposed in the literature for trajectory analysis, with applications ranging from video surveillance to sports video analysis to wildlife study. Basically, these approaches can be divided into two main categories:

- Automatic recognition of common/uncommon events, e.g. by clustering trajectories and learning motion patterns.

- Recursive composition of events, from low-level motion information to high-level scenarios.

The work presented in this study belongs to first category (due to the fact that behaviour (event) specification for fish is a complex task, because of being hardly recognizable) which is capable of automatically extracting common patterns by applying well-known data clustering methods, which minimize the effort required to define behaviours and events. Certainly, these approaches are very sensitive to how trajectories are represented and how the learning algorithm (similarity metrics as well) is tuned (for example, the number of clusters can be a fundamental choice for the success or failure of the approach). A detailed exposition of trajectory representation techniques and a comparison of similarity metrics were composed by [13]. In this work, the authors presented trajectory pre-processing and representation as two categories which involve normalization and dimensionality reduction (vector quantization, polynomial fitting, multi-resolution decomposition, Hidden Markov Models (HMMs), subspace methods, spectral methods and kernel methods). To overcome inequality between trajectory lengths zero-padding, track-extension, re-sampling and smoothing methods are suggested.

One of the earliest works on trajectory clustering was developed in [7]. In this study, object trajectories are used as a sequence of flow vectors which contains the objectŠs position and instantaneous velocity. Clustering of flow vectors was performed using a competitive neural network as level one. The output of this level is used as an input to a layer of leaky neurons to encode the partial trajectories. The partial trajectories are then used as an input of another competitive neural network. By doing this, not only clusters of normal movements but also partial trajectories were used. A similar approach to [7] was proposed by [14] using Self Organizing Maps (SOM) to detect unusual trajectories and it works on a point by point basis which makes it applicable for partially complete trajectories. Each point in the trajectory is translated into a feature vector which contains time smoothed position, instantaneous velocity and second order information, considering the short term history of motion. To find out the unusual vectors, the Euclidean distance is examined and if the distance is above a threshold then the test vector is classified as unusual.

In [26], the authors apply a grammar rule induction framework to learn event rules. A clustering approach based on [12] is used to identify simple motion patterns. Hidden Markov Models (HMMs) are trained to model each cluster, and are used as detectors of primitive events. A grammar induction algorithm, where grammars are evaluated according to the Minimum Description Length (MDL) principle [20], is then applied to build the set of event rules.

Porikli *et al.* [17] propose a method for the detection of unusual events based on spectral clustering. Histograms and HMMs based on objects' speed, color, size, aspect ratio, etc are used as features for trajectory description. For each feature, an affinity matrix (where the (i, j)-th element shows how similar the i-th and j-th objects are, according to that feature) is built and then decomposed using a certain number of the largest eigenvalues. After further transformations, a correlation matrix is computed, and clustering consists in grouping the elements which result highly correlated.

Another approach similar to [17] is presented in [25]. Each trajectory is modeled by a HMM, and a distance matrix be-

[1] *http://www.fish4knowledge.eu*

tween all training trajectories is built. Multi-Dimensional Scaling (MDS) is applied to project trajectories onto a low-dimensional space. The projected vectors are then clustered using k-means, and each cluster's trajectories are used to train a HMM representing the whole trajectory pattern. This method allows to both detect anomalous trajectories within the training set and to perform online evaluation of new trajectories by computing the matching likelihood with the cluster HMMs.

In [15], a clustering method for trajectories is presented, which can be applied both to improve tracking performance (by predicting the position of an object at time $t+1$ according to the best-matching cluster at time t) and to detect anomalous trajectories (by evaluating how frequently each cluster is matched, and considering clusters with few elements as "anomalous"). In this approach, clusters actually represent relatively short segments of trajectories, so each trajectory can be made up by segments belonging to different clusters, organized in a tree structure.

Other approaches are based on a semantic reconstruction of the scene, by recognizing just simple events at first (such as "being still" or "moving in a certain direction"), and then combining them, spatially and/or temporally, into more complex events (for example, "approaching", "following", or a combination of simultaneous sub-events). Unlike the automatic approaches, it is necessary to explicitly define rules for the description of scenarios, i.e. these algorithms are manually tuned by the users according to the scenarios they deal with and this, of course, limits their applicability. For example, Medioni *et al.*, in [11], use trajectory data and a-priori information on the scene to define three abstraction levels in the event recognition process: image features (size, speed, position, distance from reference objects), mobile object properties (entering a certain area, approaching reference objects or other actors, etc), scenarios (combinations of mobile object properties or, recursively, other scenarios). Similarly, Cupillard *et al.* in [3] model the different scenarios with "basic properties" (trajectory, speed, etc), states (a situation which involves a set of actors at a certain time, or which holds for a certain period) and events (variations of states).

3. TRAJECTORY EXTRACTION AND ANALYSIS

3.1 Fish detection and tracking

The first stage of our method consists in gathering the fish trajectories which will be fed to the clustering algorithm. The trajectory task consists of the following subtasks:

- Object detection,which, given each frame in the video sequence, detects all moving objects in the scene. The typical approach [24, 19, 16] consists in building a background model, which can be compared to the current frame to see which regions correspond to objects which do not belong to the background. Of course, in order to keep up to date with possible changes in the scene, a model update mechanism is necessary to modify the background paramters so that they match the new scene conditions. In this work we used the approach proposed in [1] which keeps for each pixel a set of its most recent colour values, against which the current pixel value is compared to check if it is a background or foreground pixel.

- Object tracking which follows an object through a sequence of frames, dealing with all related problems, such as object-object or object-background occlusions. Given the importance of this task, a lot of effort has been put by the research community into the development of accurate and robust tracking algorithms (for example, [6, 2, 18]). A common approach to tracking consists in defining a motion and appearance model for each tracked object, which allows to find the best match among the objects obtained by the detection algorithm. In this work, we adopted an algorithm specifically thought and tested for tracking fish in underwater environment, described in [23]. This algorithm uses a covariance representation of colour, texture and position features which makes it particularly suitable for tracking non-rigid objects in a noisy environment.

- Trajectory filtering: in order to remove trajectories which are likely to be caused by detection/tracking errors (for example, if a moving plant is mistaken for a fish), all trajectories are evaluated using the following approach: each association between two consecutive appearances of an object is assigned a score (obtained as a Bayesian combination a several appearance and motion regularity features, such as shape ratio, area, velocity, direction, Gabor filters, colour histograms), which reflects the likelihood that the tracking decision be correct [22]. The average value of each of these scores for a given trajectory is computed, and all trajectories which obtained a score lower than 50% are ignored. A second filtering stage is performed at the user's discretion, consisting in selecting for the clustering only trajectories having length (in terms of number of points, i.e. number of frames in which the fish appeared) included in a certain range.

- Trajectory representation: The most intuitive way, which is also the one we use in this work, is to represent a trajectory as the sequence of locations (e.g. center of the object's bounding box, object's contour mass center, etc) of the objects in each frame. Our choice is due to the fact that in our case we are particularly interested in the areas of the scene where fish tend to stay; however, in some applications, representations which focus more on the pattern of motion (such as histograms or Hidden Markov Models) rather than the sequence of locations can be more appropriate (e.g. [17, 25]).

3.2 Fish Trajectory Clustering

Once the final trajectory set has been computed, a normalization step is required, so that all trajectories have the same number of points, which is a necessary condition for the metric we adopted in the k-means algorithm. In detail, we used the Douglas-Peucker algorithm [4] to select the subset of points in a trajectory which best approximates the original curve. After that, the k-means algorithm [10] is applied to the normalised trajectories. The distance function between two trajectories of length N is defined as:

$$d(t_1, t_1) = \frac{1}{N} \sum_{i=1}^{N} norm\left(t_1\left(i\right), t_2\left(i\right)\right) \quad (1)$$

Event	*Num. trajectories*
Free swimming	2339
Sudden direction change/diving	18
Biting at coral	11

Table 1: Ground truth events and the corresponding number of trajectories.

where $t_1(i)$ is the i-th point of trajectory t_1 (similarly for t_2) and $norm(\cdot)$ is the Euclidean distance between the two points. In other words, this distance is the average distance between corresponding points of the two trajectories. Fig. 1 and 2 show an example of, respectively, fish trajectories and the clusters (with related centroids) the trajectories were clustered into.

4. EXPERIMENTAL RESULTS

In order to test the validity of the event clustering approach described above, we built a ground truth of fish trajectories corresponding to 3 different behaviours: free swimming (i.e. no unusual event), sudden direction change/diving, biting at coral.

For our tests, we set the number of clusters to 3, in order to see how well they matched the ground-truth trajectory groups. However, an advantage of our approach is that, when the number of clusters is higher than the number of ground-truth event classes, or when no ground truth exists, it allows to identify groups of trajectories which stand out by themselves, without matching a predefined behaviour, thus producing data for new possible behaviours, which will have of course to be validated by biologists.

In our experiments, we used a set of 265 videos to extract the trajectories. These videos, as we said earlier, were taken from underwater cameras located in the Taiwanese coral reef; the video resolution is 320×240 pixels, at 5 frames per second. We manually labeled 2368 ground truth trajectories, grouped between the 3 event categories as shown in Table 4.

The trajectories were normalized with the Douglas-Peucker algorithm, by resampling each trajectory to 9 points, which is the average number of points computed from all trajectories.

Table 4 shows the distribution of the ground-truth trajectories among the 3 k-means clusters.

As can be seen, the clusters are somehow able to reflect the three groups of events, although a few trajectories from the "free swimming" category fall into the second and the third cluster, which on the other hand capture with a good accuracy the rare trajectories belonging respectively to the "sudden direction change/diving" and "biting at coral".

The misclassification of the "free swimming" trajectories can be due to two main reasons: firstly, as we said in the previous sections, the manual labeling of trajectories is necessarily error-prone, since the judgement of whether a fish is performing a certain action is purely subjective; secondly, it is possible that some of the trajectories actually represent a seperate event (aside from the 3 already taken into consideration), which was not annotated in the ground truth and does not specifically fall into any of the original events.

5. CONCLUSIONS

In computer vision, event detection is a basic task for the semantic analysis of a video. In this work, we addressed the problem of detecting specific behaviours of fish in underwater real-life videos, which represent much more difficult targets than people, who are the typical subjects of the current research in the field.

The method presented in this work applies k-means clustering to a set of trajectories extracted from videos taken from Taiwan's coral reef, which have been manually labeled and categorized into 3 event groups, namely "free swimming", "sudden direction change/diving" and "biting at coral".

The results of the clustering show that this approach is able to reflect the separation between the 3 trajectory groups, although a little misclassification, which might be due to trajectory representation but most probably due to errors in the manual labeling (since fish behaviour is difficult to identify, even by watching the videos directly) or to the presence of undocumented/unlabeled behaviour whose motion patterns differ sensibly from those identified by the human operator. In fact, this is actually one of the advantages of this method, since it also allows to detect clusters of trajectories representing new data to be studied by marine biologists.

Our future efforts will thus be focused on the application of this approach to analyze videos and provide evidence of new behaviours, as well as new methods to represent trajectories (such as B-splines) and analyze trajectories (for example, by applying Hidden Markov Models to model each cluster and to evaluate a trajectory's likelihood to match it).

6. REFERENCES

[1] O. Barnich and M. Van Droogenbroeck. ViBe: a universal background subtraction algorithm for video sequences. *IEEE Transactions on Image processing*, 20(6):1709–1724, June 2011.

[2] G. R. Bradski. Computer Vision Face Tracking For Use in a Perceptual User Interface, 1998.

[3] F. Cupillard, A. Avanzi, F. Bremond, and M. Thonnat. Video understanding for metro surveillance. In *IEEE International Conference on Networking Sensing and Control 2004*, volume 1, pages 186–191. Ieee, 2004.

[4] D. H. Douglas and T. K. Peucker. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. *Cartographica The International Journal for Geographic Information and Geovisualization*, 10(2):112–122, 1973.

[5] A. Faro, D. Giordano, and C.Spampinato. Soft-Computing Agents Processing Webcam Images to Optimize Metropolitan Traffic Systems. *Computational Imaging and Vision Computational Imaging and Vision - Computer Vision and Graphics*, 32:968–974, 2006.

[6] M. Isard and A. Blake. CONDENSATION - conditional density propagation for visual tracking. *International Journal of Computer Vision*, 29(1), 1998.

[7] N. Johnson and D. Hogg. Learning the distribution of object trajectories for event recognition. In *Proceedings of the 6th British conference on Machine*

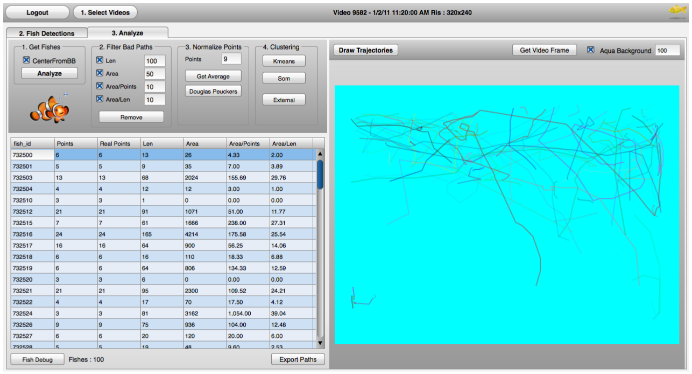

Figure 1: Example of Fish Trajectories

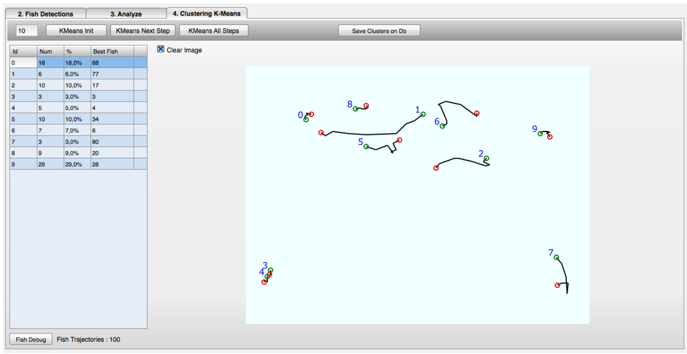

Figure 2: Centroids of the clusters the fish trajectories of Fig. 1 were grouped into.

vision (Vol. 2), BMVC '95, pages 583–592, Surrey, UK, UK, 1995. BMVA Press.

[8] B. Li and M. Ibrahim Sezan. Event detection and summarization in sports video. *Proceedings IEEE Workshop on Content-Based Access of Image and Video Libraries CBAIVL 2001*, pages 132–138, 2001.

[9] M.-Y. Liao, D.-Y. Chen, C.-W. Sua, and H.-R. Tyan. Real-time event detection and its application to surveillance systems. In *Circuits and Systems, 2006. ISCAS 2006. Proceedings. 2006 IEEE International Symposium on*, 2006.

[10] J. B. MacQueen. Some methods for classification and analysis of multivariate observations. In L. M. L. Cam and J. Neyman, editors, *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 281–297. California, USA, University of California Press, 1967.

[11] G. Medioni, I. Cohen, F. Bremond, S. Hongeng, and R. Nevatia. Event detection and analysis from video streams. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(8):873–889, 2001.

[12] M. Meila and J. Shi. A random walks view of spectral segmentation. In *AI and Statistics AISTATS*, volume 2001, pages 8–11. AISTATS, Citeseer, 2001.

[13] B. T. Morris and M. M. Trivedi. A survey of vision-based trajectory learning and analysis for surveillance. *IEEE Trans. Cir. and Sys. for Video Technol.*, 18(8):1114–1127, Aug. 2008.

[14] J. Owens and A. Hunter. Application of the self-organizing map to trajectory classification. In *Proceedings of the Third IEEE International Workshop on Visual Surveillance (VS'2000)*, VS '00, pages 77–, Washington, DC, USA, 2000. IEEE Computer Society.

[15] C. Piciarelli, G. L. Foresti, and L. Snidaro. Trajectory clustering and its applications for video surveillance. *Advanced Video and Signal Based Surveillance, IEEE Conference on*, 2005.

[16] F. Porikli. Multiplicative Background-Foreground Estimation Under Uncontrolled Illumination Using Intrinsic Images. In *in Proc. of IEEE Motion Multi-Workshop*, 2005.

[17] F. Porikli and T. Haga. Event Detection by Eigenvector Decomposition Using Object and Frame Features. *2004 Conference on Computer Vision and Pattern Recognition Workshop*, 2004.

[18] F. Porikli, O. Tuzel, and P. Meer. Covariance Tracking using Model Update Based on Lie Algebra. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 2005.

[19] F. Porikli and C. Wren. Change detection by frequency decomposition: Wave-back. In *Proc. of Workshop on Image Analysis for Multimedia Interactive Services*, 2005.

Cluster ID	*Trajectory distribution between events*		
	Free swimming	Sudden direction change/diving	Biting at coral
1	2248	4	0
2	95	14	2
3	56	1	9

Table 2: Cluster distribution of the trajectories for each ground-truth event type.

[20] J. Rissanen. *Stochastic Complexity in Statistical Inquiry*, volume 15 of *Series in computer science ; vol. 15*. World Scientific, 1989.

[21] C. Spampinato, Y.-H. Chen-Burger, G. Nadarajan, and R. B. Fisher. Detecting, Tracking and Counting Fish in Low Quality Unconstrained Underwater Videos. In *VISAPP (2)*, pages 514–519, 2008.

[22] C. Spampinato and S. Palazzo. Evaluation of Tracking Algorithm Performance without Ground-Truth Data. In *IEEE International Conference on Image Processing, to appear*, 2012.

[23] C. Spampinato, S. Palazzo, D. Giordano, F. P. Lin, and Y. T. Lin. Covariance-based fish tracking in real-life underwater environment. In *Proceedings of the International Conference on Computer Vision Theory and Applications (VISAPP 2012)*, 2012.

[24] C. Stauffer and W. E. L. Grimson. Adaptive background mixture models for real-time tracking. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, 2:246–252, 1999.

[25] N. Suzuki, K. Hirasawa, K. Tanaka, Y. Kobayashi, Y. Sato, and Y. Fujino. Learning motion patterns and anomaly detection by Human trajectory analysis. In *Systems, Man and Cybernetics, 2007. ISIC. IEEE International Conference on*, pages 498–503. Ieee, 2007.

[26] Z. Zhang, K. Huang, T. Tan, and L. Wang. Trajectory Series Analysis based Event Rule Induction for Visual Surveillance. *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, June 2007.