OUTLIER REMOVAL FOR PLAYER IDENTIFICATION IN INTERACTIVE SPORT SCENES USING REGION ANALYSIS

E. L. Andrade, J. C. Woods and M. Ghanbari

Department of Electronics Systems Engineering, University of Essex
Wivenhoe Park, Colchester, CO4 3SQ
United Kingdom
{elanet, woodjt, ghan@essex.ac.uk}

ABSTRACT

This paper demonstrates the use of segmentation, tracking, and number recognition of sport players for the provision of augmented information in moving sequences. Using prior knowledge of colour distribution the players and associated numbers are extracted for characterisation. The optical character recognition algorithms exploit multiple instances of the same character applying temporal filters to enhance reliability. Temporal analysis of the segmentation statistics provides a strong cue to the reliability of number isolation, and confirms data integrity prior to classification. Reliable segmentation, tracking and number recognition for varying scale, orientation, and motion are shown.

1. INTRODUCTION

Sports events are popular media attractions in the world today with a huge potential market impact for augmented services. The principal difficulty encountered in providing them for video is the ability of machine vision techniques to interpret the scene. Whilst extraction of generic objects from a scene is difficult, the identification of arbitrary regions based on size, colour, intensity, texture and neighbourhood is more tractable making it suitable for tracking [1].

Pre-segmentation methods [2] such as the watershed transform [3][4], and clustering [5], can result in over segmented graphs, and efficient merging algorithms are required to join adjacent and similar regions, which can be time consuming. A region adjacency graph [6] (RAG), provides a simple-connectivity view of the image and facilitates merging.

The adjacent regions can be merged provided they have sufficiently close colour distribution. Picture trees can be formed by successively merging regions against a variable threshold. These trees can help to identify isolated regions due to their reluctance to merge, and semantic objects will often be represented as a single branch of a tree

By the introduction of prior knowledge, object descriptions may be searched for in the tree structures, and therefore isolated. This is done for players in football scenes using knowledge of the clothing colour. The isolated player is then more closely scrutinised and the region containing the number is extracted. Optical Character Recognition (OCR) techniques [7] are applied to the candidate number regions, and from it the player identity derived. The recognised number can be indexed in a look-up table and used for the provision of augmented information. Reported results show that if the number can be subjectively determined, then the classifier can also. Some difficulties are experienced in the presence of rotation, lighting variation and occlusion, but the uniqueness of this application means that multiple instances of the same observation are available. This allows the application of temporal analysis, both to increase recognition count, and to remove outliers prior to classification.

2. NUMBER EXTRACTION

Despite segmentation of semantic objects being ill posed, a small amount of knowledge of a target object can considerably simplify the problem. The relational nature of regions with an appropriate description is a more robust cue than texture and colour information alone [1]. Given a pre-segmentation into a collection of homogenous regions, the problem of successfully segmenting objects becomes that of providing adequate description.

Descriptor based segmentation is seen in Fig. 1 where the sequences illustrate the isolation of the players' shirts and number extraction. Potentially insufficient resolution, blurring and occlusion may cause the numbers to be unrecognizable by the algorithm or indeed a human subject. As a general rule, if a viewer can subjectively resolve a number, then so can the recognition algorithm. A white character appended to red or black regions is searched for directly from region space using simple descriptors.

Merging is applied according to the region adjacency graph (RAG) for all candidate regions. A bounding box is framed around each candidate region set, see Fig. 1, with outliers removed according to size and aspect ratio. The resulting regions require normalisation before input to a classifier trained to detect alphanumeric characters.

Fig. 1: Illustration of the automatic number extraction algorithm. Bounding boxes are framed around number regions.

3. NUMBER RECOGNITION

There are many robust alphanumeric OCR algorithms developed for handwriting recognition e.g. for sorting letters [7]. These methods rely on accurate registration of the candidate characters i.e. framing and normalisation as seen in the previous section. This set of experiments uses the PCA classification algorithm [8]. Generally OCR is applied to a single image only, but in this work multiple instances of the same character exist across a number of frames. This fact is exploited by our algorithms. The differing viewpoints over time provide many opportunities, and time evolving analysis of classifier output indicates certainty of recognition.

The PCA classifier is applied to normalised candidate regions, and trained with ten example images of grey scale numbers with a font bearing only a close resemblance to those contained in the sequence. The minimum distance criterion that minimises the distance between the weight vector of a new image and the weight vectors of the original predefined classes is used for the classification. The recognition count for each number class is recorded. The winning class counter is incremented and the others decremented at each classification. The recognized number is the one with the highest recognition count after update. This analysis introduces temporal information in the classification process.

4. SEGMENTATION TEMPORAL ANALYSIS

The number extractor sometimes produces false candidate regions. This is due to a number of factors; including poor framing, resolution, rotation, shear, and background alias. This considered, there are two approaches that can be adopted, (i) the algorithm extracting the regions is improved, and (ii) the poor candidates are filtered from subsequent processing. Whilst (i) is undergoing constant evaluation, the focus of this section is (ii).

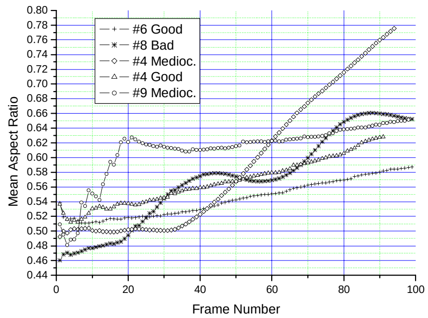
Recognition count implies the success of the classification process, but clearly this would be improved if the candidates shown to the classifier are filtered to remove outliers. To quantify an outlier from a good candidate, a number of cues are apparent e.g. the aspect ratio around the segmented number, the area and the position. Further analysis applied on a frame by frame basis, permits the parameters to be tested for temporal consistency. Clearly if consistency across a collection of frames is not observed either poor candidates are present, or a scene change has occurred. Scene changes tend to manifest as failure and discontinuities in all parameters.

The combined performance of outlier removal and temporal analysis for OCR is examined in a series of experiments. Five CIF football test sequences each one containing 100 frames are analysed. Three are outdoors (‘6 Good’, ‘4 Good’ and ‘9 Mediocre’) and show players going away from the camera, and illustrate scale dependency of OCR performance. Two sequences are indoors (‘4 Mediocre’ and ‘8 Bad’) and show a player bending from side to side to evaluate the OCR performance for rotation, shear and occlusion. Using region analysis the numbers are tracked and extracted with the appropriate descriptors, e.g. white bounded by red or black, as seen in Fig. 1. These candidate number regions are presented to the PCA classifier.

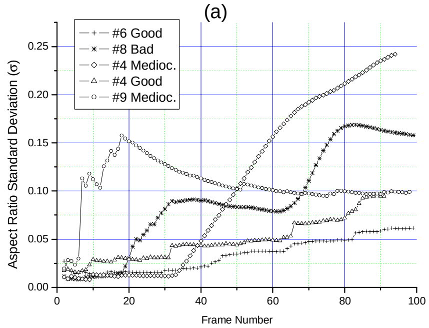
The outlier removal problem is that of detecting non-linear variations in the observed parameters. The simplest case would be to ensure the parameters’ partial differentials remain inside a given threshold. To examine this, aspect ratio, see Fig. 2, and area, see Fig. 3, are plotted for the test sequences. The sequences are subjectively evaluated as good, mediocre and bad based on the quality and visibility of the numbers. Temporal analysis of these region variables can provide temporal resilience to the classification process. For example in the test sequence the variations in the statistics of aspect ratio and area for the good sequences are notably smoother than that of the mediocre and bad sequences providing clues for the tracking reliability.

The effect of outlier removal prior to classification is analysed for the test sequences. An observation at time t, $O(t)$ is considered an outlier if it differs from the current mean value, $\mu(O(t))$, of the series by more than α times the current series standard deviation $\sigma(O(t))$, as defined in equation (1).

$$\left|O(t)-\mu(O(t))\right|\geq\alpha\sigma(O(t))\rightarrow Outlier \tag{1}$$



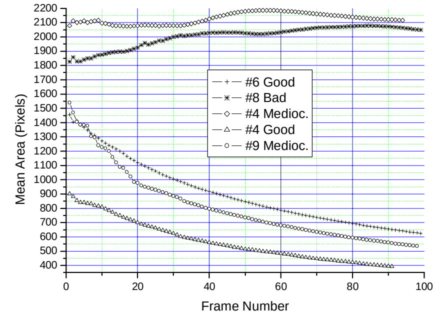
(a)



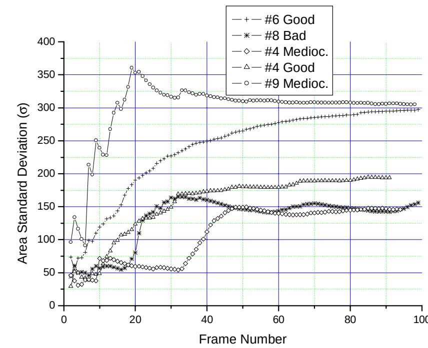
(b)

Fig. 2: Statistics of aspect ratio for the test sequences; (a) mean and (b) standard deviation.

Two parameters are evaluated for outlier removal aspect ratio and area. For each one α is varied from 0.1 to 6.0 standard deviations in the decision criterion. The filtered sequences are presented to the PCA classifier and the recognition count is updated accordingly. The analysis look for the value of α, which provides the minimum classification error, and simultaneously gives the best recognition performance. Table 1 shows the best values of α according to these criteria, the ratios of correct and incorrect classification, the percentage of non-filtered candidate regions that is processed, and also the subjective classification of the test sequence in the leftmost column. Table 2 shows the same variables for the area parameter. The 'Good' sequences have an excellent classification performance independent of which variable is observed the 'Mediocre' sequences have an improvement when aspect ratio is used as the filtering parameter instead of area. The 'Bad' sequences do not benefit of the filtering due to the poor quality of the extracted candidate regions. Area is a conservative parameter for outlier definition since it rejects most of the samples due to high variability in 'Bad' and 'Mediocre' sequences.



(a)



(b)

Fig. 3: Statistics of area ratio for the test sequences; (a) mean and (b) standard deviation.

Aspect Ratio – Criterion: Minimum Classification Error Ratio					
Quality	Number	α	Correct Ratio	Incorrect Ratio	% of Samples
Medioc.	4 – rot.	1.5	0.27	0.01	36.17
Bad	8 – rot.	0.1	0.02	0.04	06.06
Good	6 - scal.	3.6	0.97	0.01	98.99
Medioc.	9 –scal.	1.7	0.89	0.03	91.83
Good	4 –scal.	4.0	0.98	0.01	98.90

Table 1: Best classification performance for minimum error criterion with outlier removal based on aspect ratio for all test sequences.

Area – Criterion: Minimum Classification Error Ratio					
Quality	Number	α	Correct Ratio	Incorrect Ratio	% of Samples
Medioc.	4 – rot.	0.1	0.05	0.02	7.44
Bad	8 – rot.	0.1	0.02	0.05	7.07
Good	6 - scal.	1.9	0.98	0.01	100
Medioc.	9 –scal.	0.1	0.89	0.03	2.04
Good	4 –scal.	2.2	0.98	0.01	100

Table 2: Best classification performance for minimum error criterion with outlier removal based on area for all test sequences.

The improvement of outlier removal in a 'Mediocre' sequence is observed comparing the recognition count without outlier removal Fig. 4 for the sequence '9', against Fig. 5, where outliers are removed using the best α shown in Table 1. The segmentation errors occurring in the first 30 frames, misclassifying number 9 as number 6 are filtered out, thus not affecting the temporal analysis of the recognition count as a classification criterion.

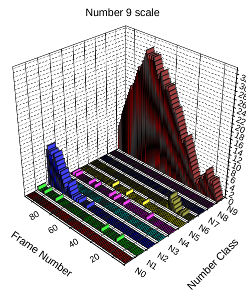


Fig. 4: Recognition count for test sequence ‘9’.

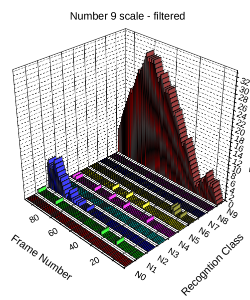


Fig. 5: Recognition count for test sequence ‘9’ with outlier removal.

5. CONCLUSION

In this paper, we have demonstrated the use of segmentation, tracking, and number recognition of sport players for the provision of augmented information in moving sequences. A combination of region space analysis, database retrieval, prior knowledge, relational descriptors and optical character recognition, have allowed this in a reliable and quantifiable way. With tracking dependant on the segmentation, and segmentation identified as ill defined, simple prior knowledge about the players is introduced to highlight the required objects. Segmented players are subjected to closer scrutiny using the same techniques to identify the number regions. These segmentations were shown not to always be reliable, containing partial, or even null observations of the target numbers. Candidate regions are subjected to normalisation making them suitable for the OCR algorithm. For appropriately extracted segmentations, the recognition is very reliable, but often candidates cannot be recognised either by human observation or by the algorithm. Occasional error is permissible since by virtue of tracking, multiple instances of the same character are available for classification across a number of frames. This is exploited by the use of temporal analysis. Temporal filtering and accumulators are used to improve the recognition count by reducing the impact of misclassifications, and improving the visualisation. Temporal reliability measures can be derived by examining the transitions of the variables associated with the candidate regions. These include aspect ratio, area, and recognition count. Aspect ratio and area are strongly correlated with the recognition performance, but are susceptible to rotation. The inter-relational nature of these two variables can provide cues to the type of motion present e.g. object going into the distance. Analysis of the temporal behaviour of these variables permits the removal of outlying candidate regions for classification thereby providing number recognition with temporal derived resilience. Currently a dynamical model for the parameters and recognition performance is under development to provide a quality metric for segmentation.

6. ACKNOWLEDGEMENT

This research is supported by the European Union project MELISA.

7. REFERENCES

[1] E. L. Andrade, E. Khan, J. C. Woods, M. Ghanbari, “Description based object tracking in region space using prior information”, Electronic Letters, vol. 39, no. 7, April 2003.

[2] P. Salembier and F. Marqués, “Region-Based Representations of Image and Video: Segmentation Tools for Multimedia Services”, IEEE Transactions on Circuits and Systems for Video Technologies, vol. 9, no. 8, pp 1147-1169, December 1999.

[3] L. Vincent and P. Soille, “Watersheds in Digital Spaces: An Efficient Algorithm Based on Immersion Simulations”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 13, no.6, pp 583-598, June 1991.

[4] I. Vanhamel, I. Pratikakis, H. Sahli, “Multiscale Gradient Watershed of Color Images”, IEEE Transactions on Image Processing, vol. 12, no. 6, pp 617-626, June 2003.

[5] I. Kompatsiaris, E. Triantafillou, and M. G. Strintzis, “Region-based color image indexing and retrieval”, Proceedings of the International Conference on Image Processing, vol. 1, pp 658 -661, October 2001.

[6] Y. Tsaig and A. Averbuch, “Automatic Segmentation of Moving Objects in Video Sequences: A Region Labeling Approach”, IEEE Transactions on Circuits and Systems for Video Technology, vol. 12, no. 7, pp 597-612, July 2002

[7] R. Plamondon and S. N. Srihari, “On-Line and Off-Line Handwriting Recognition: A Comprehensive Survey”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no. 1, pp. 63-84, January 2000.

[8] M. Turk and A. Pentland, “Eigenfaces for Recognition”, Journal of Cognitive Neuroscience, vol. 3, no. 1, pp. 71-86, March 1991.