Pseudo Fuzzy Colour Calibration for Sport Video Segmentation

Catarina B. Santiago, Armando Sousa
Faculty of Engineering of the University of Porto, Porto, Portugal
Institute for Systems and Computer Engineering of Porto (INESC Porto), Porto, Portugal
Luis Paulo Reis
Faculty of Engineering of the University of Porto, Porto, Portugal
Artificial Intelligence and Computer Science Laboratory (LIACC), FEUP, Porto, Portugal

Video segmentation is one of the most important parts of a vision system which allows partitioning each frame into homogeneous regions that share a common property. This work proposes a new methodology that aggregates three different techniques: background subtraction, region growing and a pseudo Fuzzy colour model to define colour subspaces that characterize each class. In addition, the pseudo Fuzzy colour model allows a given colour to belong to more than one class and enables the expansion of the classes through a dynamic model based on belonging and persistence information. In case of shared colours among classes, regional features are searched in order to determine the object's class. Tests with test and real videos of sports footages show promising results.

KEYWORDS: Computer Vision, Video Segmentation, Fuzzy Logic, Sports Videos

1 INTRODUCTION

Video and inherently image segmentation is the first step and probably the most critical step in any vision system. In fact, the quality of the final result is highly dependent on a good segmentation.

In this paper we present a methodology that combines colour region growing with a pseudo Fuzzy pixel labelling methodology to subdivide the colour space into colour classes, that may or may not be disjunctive depending on the objects to be segmented.

The methodology followed includes the identification of foreground pixels using background subtraction, the definition of colour classes for each object or groups of objects using a pseudo Fuzzy model, the labelling of foreground pixels into the corresponding object and a continuously update on the colour classes that characterize each object or group of objects. Preliminary results using test and real sport images show that this is a promising technique that may be used for sport video segmentation in order to take into account light variations among frame regions (due to shadows of other objects or non-uniform illumination) and between frames.

The paper structure is as follows. The next section presents some of the most used colour segmentation methodologies, Section 3 describes the algorithm implemented for performing the video segmentation. Section 4 presents the results achieved and finally section 5 concludes this paper presenting and draws some future work directions.

2 RELATED RESEARCH

Video segmentation can be seen of different perspectives, one may want to segment the video into meaningful temporal sequences which is known as temporal segmentation. Temporal segmentation usually corresponds to the first step of video annotation and tries to segment the video taking into account similarities/dissimilarities between successive frames (Koprinska and Carrato 2001).

On the other hand, one may be interested on analysing the content of each frame and extract information concerning the objects that are present in it, or in other words divide each frame content into homogeneous regions that correspond to independent objects and therefore perform a spatial segmentation. Despite the name, spatial segmentation, may be accomplished using temporal characteristics of the video as will be shown latter on this section.

The focus of this work is more on spatial segmentation, therefore for a detailed survey on temporal video segmentation please refer to (Koprinska and Carrato 2001).

Spatial video segmentation may be performed using the methodologies that are used for image segmentation and further enhanced using the temporal characteristics of video. In addition, when performing colour analysis there is also the need to choose a colour space.

Regarding colour image segmentation, a detailed survey is given by (Cheng et al. 2001). As they state, most of the existing colour image segmentation methodologies have their origins on grey scale image segmentation with the addition of a proper colour space choice.

There have been proposed several colour spaces. (Vandenbroucke et al. 2003) provide a taxonomical classification of colour spaces into primary colour spaces (RGB, XYZ) which result from the trichromatic theory, luminance-chrominance spaces (YIQ, YUV, L*a*b*), perceptual spaces (HSI, HSV, LCH) that try to mimic the human perception regarding to colour and independent axis spaces ($I_1 I_2 I_3$).

Additionally, they propose a new hybrid colour space that chooses the colour components of different colour spaces that best characterize each pixel class.

The main categories of image segmentation methodologies (Cheng et al. 2001) include:

- histogram thresholding by determining the peaks or modes of the multi-dimensional histogram of a colour image
- feature space clustering by grouping the image feature space into a set of meaningful groups or classes based on intensity, colour or texture characteristics of pixels and not on the spatial relation among them
- region based which include region growing, Watershed transform and region split and merge. These methods try to divide the image domain based on the fact that adjacent pixels in a same region have similar visual features (colour, intensity, texture or motion)
- edge detection methods that segment the image by finding the edges of each region using one of the well known edge detectors
- Fuzzy methods allow classes and regions to have a certain uncertainty and ambiguity which is general the case in image processing
- neural networks are very powerful tools that allow parallel processing and the incorporation of non-linearities. They can be used either to pattern recognition, classification or clustering.

Nowadays there is the tendency to aggregate techniques from different categories in order to achieve better results. A typical example of this case is the JSEG algorithm (Deng and Manjunath 2001) that initially clusters colours into several representative classes, afterwards replaces each pixel by their corresponding colour class label and only then a region growing process is applied directly to the class map.

On videos, contrary to static images, besides the two physical components and colour information there is also the time component. Using this property it is possible to segment images based on motion along time. In order to perform this task there are two main approaches, background subtraction and optical flow.

Background subtraction can be used in cases where a more or less fixed background can be assumed and in this case it is possible to subdivide the image into foreground and background. Several background subtraction techniques have been proposed in literature. The main issues on these methods is to obtain a good estimate of the background.

The simplest method to model the background is to use a single static image without objects. However this approach works rather poorly, because it does not take into account changes that may occur in the background (for example light effects). More robust methods include estimating the background model using a moving average (Heikkila and Silvén 1999), median or even a mixture of Gaussians (Grimson et al. 1998). Having the foreground regions detected it may still be necessary to perform their labelling or categorization.

Optical flow (Barron and Thacker 2005) is based on the fact that when an object moves in front of a camera, there is a corresponding change in the image, however it assumes small displacements during time.

In this paper we propose a methodology that combines the ideas behind three of the described segmentation methods: background subtraction for detecting foreground regions, region growing and pseudo Fuzzy categorization for colour calibration.

3 PROPOSED VIDEO SEGMENTATION METHODOLOGY

The first step for the video segmentation to take place consists on a supervised calibration of the colours of each class which is achieved using a region growing method. The initial colour seeds for each class are set manually using the mouse to click on the objects that will be segmented, afterwards the surrounded pixels are agglomerate around these seeds using colour distance criteria. The colour expansion is performed on the HSL (Hue, Saturation and Luminance) colour space in order to minimize the effects of shadows and light variations.

The regions growth is performed in all directions in a recursive way until reaching a pixel that has a colour too far away from the seed or from the previous neighbour. During the colour expansion each pixel is attributed a given belonging degree to the class be-

ing calibrated (the number of classes is defined by the user). This degree is stored in a lookup table that contains for each colour triplet the belonging degree to each class. By using the lookup table it is possible to have a fast access to this information latter on the segmentation process.

The belonging degree can have four levels: no belong (by default and before the calibration takes place, all the colours are categorized with no belong degree to every class), low belong degree, full belong and additionally a full belong degree with the characteristic of also being a colour seed. The following heuristics (applied sequentially) are used to attribute the belonging degree to each colour during the region growing process:

- if the pixel was assigned to the class and physically is quite close to the initial seed pixel then it is also assumed to be a seed pixel with a full belonging degree
- if the Euclidean colour distance to the initial seed pixel is less than two thirds the maximum allowed distance for the growing then the pixel is categorized with a full belonging degree but without being a seed
- otherwise the pixel is categorized as belonging to the class with a low belong degree

By the end of the calibration process the colour space is subdivided into classes, which are not necessary disjoint since the same colour can belong to different classes with a given belonging degree.

The motivation behind having non-disjoint classes is related with the fact that objects may have different colours and also different objects may share a common colour, for example in the case of sport videos it is common that team have uniforms with white stripes.

Once the object classes have been calibrated, the image segmentation can take place. The first step consists on eliminating background regions. Since the background is more or less static on the images we want to deal with, the subtraction is performed using a blank image of the viewed scene and a dynamic threshold that is updated for each pixel in each frame. Only the pixels that are classified as background see their respective threshold updated, the pixels classified as foreground remain with the associated threshold unchanged. The update obeys to equation 1 and its value is never allowed to be below 4% of the entire colour range (0-255). This value was obtained experimentally.

$$\sigma^c_{t+1}(x,y) = \begin{cases} \alpha(I^c_t(x,y) - B^c(x,y)) + (1-\alpha)\sigma^c_t(x,y) \\ \qquad \text{if } I_t(x,y) \in B(x,y)) \\ \\ \sigma^c_t(x,y) \quad , otherwise \end{cases} \tag{1}$$

Where

- σ is the threshold of the pixel at position *(x,y)*, time *t+1* and colour component *c*,
- *I* is the colour intensity of the pixel at position *(x,y)*, time *t+1* and colour component *c*,
- *B* is the background colour intensity of the pixel at position *(x,y)*, time *t+1* and colour component *c*,
- α is a learning constant, that for our specific case was set to 0.02.

Pixels whose colour difference to the background image is less than the respective threshold are labelled as background, the others are labelled as foreground.

After the foreground pixels are identified, their colour is compared against the colour lookup table that resulted from the calibration process and classified into one of the classes. Since the same colour can belong to different classes it may occur that a pixel is classified into more than one class. To break this tie, information from adjacent pixels that have already been classified is used. The number of adjacent pixels that belong to each class are counted and those classes that have a count higher than one and a half times the minimum count are assigned two extra points in the belonging degree. This way, it is possible that, although the belonging degree of a pixel to a class based on the colour calibration information is lower than the belonging degree to another class it may be the winner due to the neighbourhood characteristics.

Additionally, if the winning class has a full belong to that colour triplet and corresponds to a seed colour then a region growing process is triggered and the colour lookup table that contains the information concerning the classes colours is updated. This auto expansion is more restrictive than the one performed during the manual initialization and is not performed at every frame, otherwise the processing would be too time consuming.

In order for this update to add not only colour triples to the classes but also to remove them (otherwise classes would grow too much), each colour triplet has associated a persistence to that class. Colours with lower belonging have lower persistence and colours with higher belonging have higher persistence. The initial persistence given to the colour is proportional to the time between auto expansions. This proportion factor is $1/8$ for low belong colours and $1/4$ for full belong colours.

The persistence is maximum when the colour is added to the class and diminishes whenever it is not "seen" in a frame, however seed colour have infinite persistence and will therefore always remain in the class. Whenever the persistence value reaches zero the colour triplet is removed from the class.

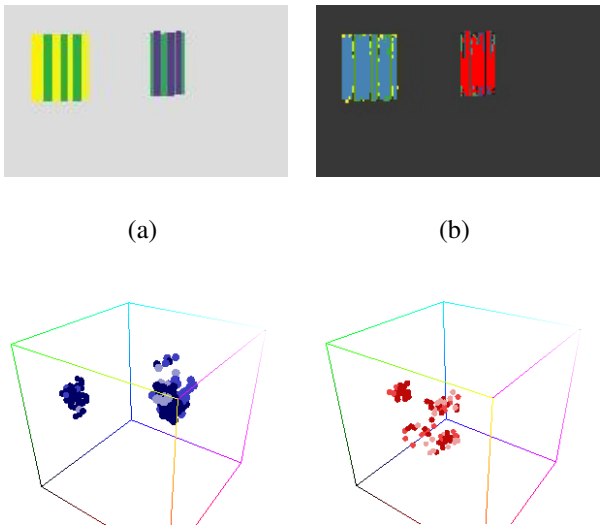
With the introduction of this dynamic it is possible to have mutable classes that adapt to light changes either occurring at different regions of the same frame or between frames.

At the same time the foreground pixels are classified, they are also aggregated horizontally to form run length encoding (RLE) structures that contain a label indicating the class they belong to, the y, x_{min} and x_{max} positions. Small RLEs are ignored in order to minimize noise. Finally the RLEs are merged vertically to form blobs. A full description of this pixel aggregation can be found in (Santiago et al. 2011).

4 RESULTS

In order to validate the proposed methodology a test video was made that consists on two colour squares with stripes as illustrated on Figure 1(a).

A few mouse clicks on the image resulted on the colour calibration of Figures 1(c) and 1(d), the different tones indicate different belonging degrees, also it is possible to verify (as expected) that the two colour classes superimpose on the green region of the global colour space.



(a) (b)

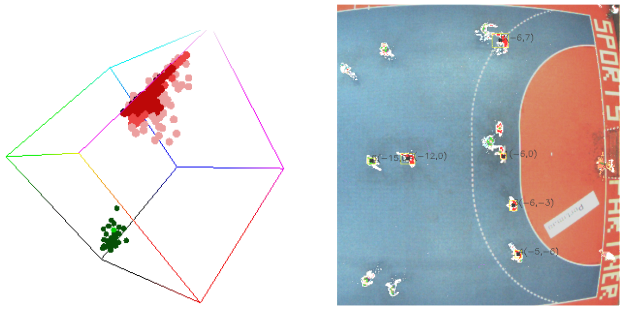
(c) (d)

Figure 1: (a) frame from the test video, (b) segmented frame, (c,d) colour classes resulting from the initial calibration for the left and right objects (different colour tones indicate different belonging degrees).

On the final segmented image (Fig. 1(b)) the background is darkened and the pixels belonging to each object are identified by the respective colour class label. It is also possible to verify that although the green colour is common to both objects the pixels are correctly labelled (due to the characteristics of the adjacent pixels).

In order to validate the continuously and automatic update of the colour classes, tests using a video of a sport match were performed. For the tests presented here the auto expansion process occurs at every 60th frame. It is important to highlight that videos with a more dynamic behaviour (which is the case of sport videos) require an update frequency higher than videos that are not so dynamic.

Initially the colour classes were calibrated using the mouse to click on the players which resulted on the colour classes of Figure 2(a) and the respective processed image of Figure 2(b).

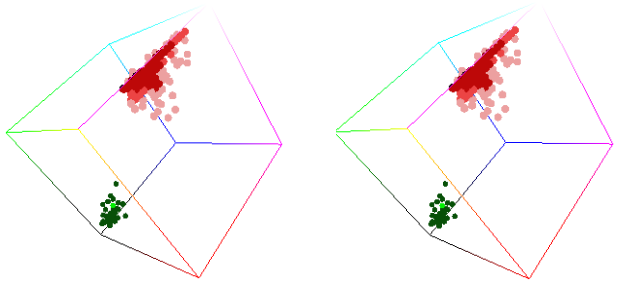


(a) (b)

Figure 2: Initial colour calibration. (a) colour classes being segmented and (b) segmented image.

After 900 frames (which correspond to 15 auto expansions) it is possible to verify that the colour classes have grown around the initial seeds (represented by the darker colours) which resulted on the updated colour classes of Figure 3(a). During this time players from both teams stayed on the area under analysis.

Afterwards, the players permanence on the area decreased during 2050 frames (34 auto expansions) which caused the colour classes to retract a little (Fig. 3(b)).



(a) (b)

Figure 3: Results of the colour auto expansion process. (a) Colours classes after 15 auto expansions. (b) Colour classes after more 34 auto expansions.

Table 1 provides an overview of the changes between these three cases. From the initial set 2(a) to Fig 3(a) set the number of colour triplets that fully

belong to the two classes increases and the pixels that have a low belong degree decrease because the auto expansion process is more restrictive than the initial manual calibration.

Regarding the evolution to the colour classes of Fig 3(b), the number of triplets that belong to the red class decrease less (compared to the time of Fig 3(a)) then the ones of the green class because the players of the first class are the last to leave the area and the first to enter it again, since they are the defending team.

Table 1: Number of colour triples that belong to each class and the respective belonging degree for the initial set 2(a), set of Fig 3(a) and set of Fig 3(b)

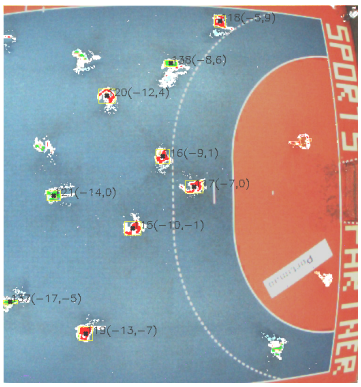
Class	Belong	Init Set	Fig 3(a)	Fig 3(b)
Green	Low	1	0	0
	Full	0	8	0
	Full+seed	33	34	34
Red	Low	43	27	26
	Full	16	18	16
	Full+seed	27	43	44

Comparing the results with and without the pseudo Fuzzy model of colour expansion it is possible to verify that the segmentation has better performance with the mutable colour classes as depicted on Fig 4. Fig 4(a) has more players from the green class detected and the players of the red class have more area detected.

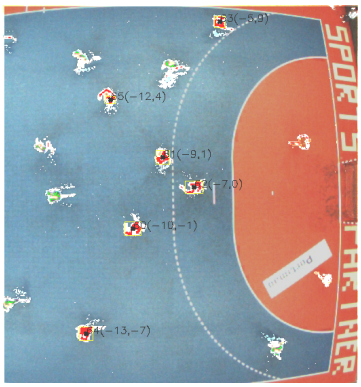
5 CONCLUSIONS

In this paper we presented a methodology for segmenting sport video images. The main objectives were to develop a methodology that allowed not only the update of the background/foreground models but also of the object colour subspaces by continuously and automatically updating the colour classes in order to take into account different light conditions. The usage of a Pseudo Fuzzy technique allowed this update (increase or decrease the number of triplets to the classes) by taking into account the belonging degree of each colour triplet to the respective class. In addition, the Pseudo Fuzzy model enabled the same colour triplet to belong to different classes. The tie break was achieved by evaluating the characteristics of the adjacent pixels.

The proposed methodology was validated using a test video but also a real video from a sport match. Results are quite promising, but more tests with real videos must be performed not only with different teams but also with teams that share a common colour in their equipments. It would also be interesting to test this approach in videos where the objects to be segmented have higher areas, since the players are quite small in the video used.



(a)



(b)

Figure 4: Result of the segmentation: (a) with colour auto expansion and (b) without colour auto expansion.

ACKNOWLEDGMENTS

We would like to thank Fundacao Calouste Gulbenkian by the support given trough a PhD scholarship with ref. 104410.

REFERENCES

Barron, J. and N. Thacker (2005). Tutorial: Computing 2D and 3D optical flow. *Tina Memo Internal* (2004-12).

Cheng, H., X. Jiang, Y. Sun, and J. Wang (2001). Color image segmentation: advances and prospects. *Pattern recognition 34*(12), 2259–2281.

Deng, Y. and B. Manjunath (2001). Unsupervised segmentation of color-texture regions in images and video. *IEEE Transactions on Pattern Analysis and Machine Intelligence 23*(8), 800–810.

Grimson, W., C. Stauffer, R. Romano, and L. Lee (1998). Using adaptive tracking to classify and monitor activities in a site. In *Computer Vision and Pattern Recognition, 1998. Proceedings. 1998 IEEE Computer Society Conference on*, pp. 22–29. IEEE.

Heikkila, J. and O. Silvén (1999). A real-time system for monitoring of cyclists and pedestrians. In *Visual Surveillance, 1999. Second IEEE Workshop on,(VS'99)*, pp. 74–81. IEEE.

Koprinska, I. and S. Carrato (2001). Temporal video segmentation: A survey. *Signal processing: Image communication 16*(5), 477–500.

Santiago, C., A. Sousa, L. Reis, and M. Estriga (2011). Real Time Colour Based Player Tracking in Indoor Sports. *Computational Vision and Medical Image Processing*, 17–35.

Vandenbroucke, N., L. Macaire, and J. Postaire (2003). Color image segmentation by pixel classification in an adapted hybrid color space. Application to soccer image analysis. *Computer Vision and Image Understanding 90*(2), 190–216.