Luis Alvarez, Pedro Henriquez, and Javier Sanchez

CTIM (Centro de Tecnologías de la Imagen), Universidad de Las Palmas de Gran Canaria, Spain {lalvarez@ctim.es,phenriquez@ctim.es,jsanchez@ctim.es} http://www.ctim.es

Abstract. In this paper we present a new method for image primitives tracking based on a CART (Classification and Regression Tree). Primitives tracking procedure uses lines and circles as primitives. We have applied the proposed method to sport event scenarios, specifically, soccer matches. We estimate CART parameters using a learning procedure based on RGB image channels. In order to illustrate its performance, it has been applied to real HD (High Definition) video sequences and some numerical experiments are shown. The quality of the primitives tracking with the decision tree is validated by the percentage error rates obtained and the comparison with other techniques as a morphological method. We also present applications of the proposed method to camera calibration and graphic object insertion in real video sequences.

Key words: CART, decision tree, vision

1 Introduction

Camera calibration of video sequences is an important issue in computer vision. Most calibration methods consist of several stages, which usually include initialization, calibration estimation, primitive tracking and calibration refining. In particular, the broadcasting of sport events increasingly introduces the processing of video sequences for a wide variety of purposes, some of them require a highly precise and fast calibration.

In this paper, we focus on the application of CART (decision tree techniques) to tackle the primitive tracking stage in a video sequence which is an important step to perform the calibration and it is a procedure which takes an important part of the processing time. We analyze this problem in real application scenarios, where we deal with some additional problems, such as the small number of primitives which are usually visible, or the large size of HD videos. The main assumption we make is, that the background and the primitives are different enough in the RGB space.

Our camera calibration implementation is an incremental procedure where the calibration of a frame is calculated with information obtained from the previous two frames. By camera calibration we mean the estimation of camera rotation and translation in the 3D space, as well as the camera intrinsic parameters (zoom setting, pixel aspect ratio and the projection of principal point in

the image). Primitives are the lines and circles in a model of a soccer field with actual dimensions, which are projected using the homography estimated from the previous two frames. It means that in the primitive tracking stage we search for the primitives in the image using the estimated projection of the reference primitives as start point. Then we use a decision tree to determine primitive location. Since we start searching the primitive from a near point, always inside the soccer field, we only need to differentiate two classes with the decision tree, primitives and grass. Instead of processing all the image pixels, as is done in [9], we only have to classify pixels in a primitive neighborhood of the previous primitive location.

This paper is organized as follows: In section 2, we summarize the state of the art. Section 3 explains the decision tree building and the learning procedure. In section 4, we introduce a method to perform primitive tracking using a decision tree. In section 5, we show some experiments and results. Finally, in section 6, we present our main conclusions.

2 Related works

Nowadays, decision trees are widely used in image segmentation and classification because they generate easy-to-understand rules and they are fast to train and apply see for instance [14] and [13]. Furthermore, decision trees can manage different features to classify the pixels. This classification is used on images in different areas, for instance, medical images [10] or satellite images [11, 12]. They usually work with RGB (Red-Green-Blue) channels combined with other color spaces, such as HSV (Hue-Saturation-Value) or the different bands of satellite images, for example infrared. In this paper, we apply a decision tree technique to perform the primitive tracking as part of a camera calibration procedure. For primitive tracking, different methods have previously been proposed. In [5], a line tracking method guided by camera parameters is described. A correlation technique is applied in [4]. In [8], camera parameter ranges are defined and a search for the optimal matching is carried out. The approaches presented in [6] and [7] use YCbCr color space to detect white lines in the image. Finally in [9], the primitives are tracked with a morphological line detection method. The method we propose uses CART (decision tree) to classify pixels into primitive pixels and background pixels based on a learning procedure on RGB image channels. A general introduction to CART is presented in [1].

3 Decision tree building and learning to classify primitives

A CART, as those described in [1], is used to detect the white primitives. One set of features is considered to discriminate different pixel classes. In the learning procedure of the decision tree, it is important to determine the way to select the channels and thresholds in order to build a simple tree.



Fig. 1. Binary decision tree. Points n_i in node T1 have a value *i* in channel X less than threshold C_i . Points n'_i in node T2 have $X_i > C_i$.

To build the decision tree is necessary a learning stage based on a training set with information about different classes, see for instance [2]. Then for each video sequence, we have a classification data set, which contains information about two classes, primitives and background. Usually, in our soccer field scenarios, primitives are white and the background is green. In the data set, there are RGB values obtained from a manual segmentation of the first frame of the sequence. These RGB triplets are taken to build a three-channel decision tree. It determines, within each node, which channel provides the best discrimination. With this purpose, we use a measure to estimate the impurity of the sets based on Gini's index for implementation simplicity, see for instance [3]:

$$\sum_{K \neq K'} P_K P_{K'} = \sum_{K=1}^N P_K (1 - P_K)$$
(1)

where P_K is the probability of a point belonging to a class. In the learning procedure of the decision tree, in order to decide the channel and the threshold which are selected within each node, we have to minimize the resulting impurity measure by dividing the set of points (see Figure 1). The aim is to find the values X_i and C_i which minimize the compound energy in Equation 2, which is the addition of the energies of both son nodes:

$$E_{t} = E_{0} \left(X_{i}, C_{i} \right) + E_{1} \left(X_{i}, C_{i} \right)$$
(2)

$$E(X_i, C_i) = \sum_{i=1}^{n} \frac{n_i}{n_1 + n_2 + \dots + n_n} \left(1 - \frac{n_i}{n_1 + n_2 + \dots + n_n} \right)$$
(3)

4 Primitive tracking by means of a decision tree

For each frame, first, we initialize the primitive location using the information of 2 previous frames. The resulting primitive points can be in two situations: either on a white primitive or on the background. If the projected point was classified into the primitive class by the decision tree, we must search for the edges of the primitive in both directions of the perpendicular orientation (see Figure 2

4 CART application to image primitives tracking



Fig. 2. Search for the edges in the primitive tracking procedure. The red point is the point initialization based on the previous frames. The blue dot-line is the orthogonal line that we examine to find the edges. The green square is the center of the primitive.

left). Once the edges have been found, we calculate the midpoint between both, which will be considered as the center of the primitive. If the projected point was classified into the background class by the decision tree, we also search for a primitive pixel in both directions of the perpendicular orientation. However, as soon as a primitive pixel is found, we only continue moving in that direction, searching for the other primitive edge. Once it has been found, we calculate the midpoint (see Figure 2 right).

In both situations, in order to avoid considering large white zones, such as advertisements or players wearing white clothes, we have to control the thickness of the primitives with a threshold. This threshold is dynamically obtained, because farther primitives are thinner than closer ones. We calculate this threshold as the distance between the projections of a reference point and another point obtained with the addition of a certain thickness. If we have examined more pixels than the width limit, the primitive detection is rejected at that point. When the lines and circles are tracked, they are used as a calibration pattern.

5 Numerical experiments and results

We have tested our method on different video sequences using both, scale soccer court models (1440 x 809) and real scenes from soccer matches in HD (1920 x 1080). The training data set is manually obtained by segmenting the first frame of the sequence, with only two different classes: primitives (white lines and circles) and grass. Examples of manual segmentation are shown in Figure 3 and Figure 4.

In order to achieve the best results for classification, we have tested different CART configurations varying the number of channels. These tests consist in build a decision tree with a set of training data. After the learning procedure, the data is classified with the decision tree and the pixels classified with errors are counted. Finally, we compare decision tree classifications with manual classifications on four random frames extracted from each video, which show different parts of the soccer field. The results are shown in Table 1 and Table 2, where the first row shows the classification results of training data set and other rows

 $\mathbf{5}$



Fig. 3. Two different classes are used in the manual segmentation: white primitives and grass. Grass is segmented using polygons, whereas segments are used for primitives. Scale model image.



Fig. 4. Two different classes are used in the manual segmentation: white primitives and grass. Grass is segmented using polygons, whereas segments are used for primitives. Real image.

are the classification results for some random video sequence images. We can also see these results in Figure 5 and Figure 6. We can observe that the best configuration for the decision tree channels is RGB, it has the lowest classification error rate. The performance of the decision tree is also illustrated in Figure 7 and Figure 8, where some images with the classification results are obtained. We note that we only focus in the grass/primitive discrimination in a neighborhood of primitive initialization so it does not matter what is the result of the decision tree for other areas in the image. We have compared the classification results obtained with decision tree using RGB and the results of classifying with the morphological method [9]. We have tested both with four manually segmented images from two different videos, scale model and real soccer match videos. In Table 3 the percentage error rates are shown, and in Figure 9 we can see the comparison between the two methods.

 $\mathbf{6}$

Table 1. Decision tree configurations percentage error rates, scale model images.

R	G	В	Η	\mathbf{S}	\mathbf{V}	RGB	HSV
0.068	0.268	0.153	1.112	0.060	0.268	0.005	0.009
0.067	0.128	0.100	0.563	0.055	0.128	0.049	0.051
0.087	0.119	0.101	0.399	0.065	0.119	0.066	0.072
0.069	0.110	0.067	0.270	0.076	0.110	0.059	0.046
0.049	0.049	0.076	0.475	0.035	0.091	0.040	0.047

Table 2. Decision tree configurations percentage error rates, real images.

R	G	В	Η	\mathbf{S}	V	RGB	HSV
0.200	0.344	0.194	0.398	0.322	0.344	0.067	0.070
0.174	0.246	0.157	0.297	0.230	0.246	0.121	0.124
0.231	0.358	0.201	0.414	0.322	0.358	0.102	0.105
0.215	0.268	0.208	0.396	0.315	0.268	0.163	0.166
0.146	0.292	0.154	0.263	0.206	0.292	0.124	0.134

An additional interesting issue in the proposed method is the processing time. The morphological method described in [9], takes an average of 2725 ms to process a HD frame in a four-core processor. The method we propose in this paper using a decision tree takes 7 ms per HD frame with the same processor. In terms of computational complexity the main difference between both methods is that a decision tree computation is very fast and the method is local (i.e. we only need to process a neighborhood of primitive location). Morphological operation takes much more time and the procedure process the whole image. In Figure 10, we show the primitive tracking stage in a frame with all the points which have been analyzed to find the white primitives.

Table 3. Comparison of percentage error rates between decision tree method and morphological method. RGBs and RGBr are results obtained with CART in scale model images and real images. Ms and Mr represents the results with morphological method.

RGBs	Ms	RGBr	\mathbf{Mr}
0.049	0.072	0.121	0.198
0.066	0.253	0.102	0.312
0.059	0.259	0.163	0.328
0.040	0.084	0.124	0.185



Fig. 5. Decision tree configurations percentage error rates, scale model images. These rates are from Table 1 $\,$



Fig. 6. Decision tree configurations percentage error rates, real images. These rates are from Table 2 $\,$

To illustrate the quality of the results, the calibration we have obtained is used to calibrate the video sequence and to insertion of graphics into the video. With the camera parameters obtained from the calibration, we can synchronize the real camera with a virtual camera and render objects with the same perspective. We have chosen this kind of experiments because these applications require an accurate and fast video calibration estimation. CART primitive tracking method is fast and precise detecting the line centres. The results are shown in Figure 11.

Finally, in http://www.ctim.es/demo105/ we show some video sequences used in the experiments and classified with the CART.

8 CART application to image primitives tracking



Fig. 7. Scale model image segmented by means of the CART method. The points classified as primitive point are painted in yellow. Videos in http://www.ctim.es/demo105/



Fig. 8. Real image segmented by means of the CART method. The points classified as primitive point are painted in yellow. Videos in http://www.ctim.es/demo105/



Fig. 9. Comparison of percentage error rates between decision tree method and morphological method on scale model(left) and real images(right). These rates are from Table 3



Fig. 10. Primitive tracking stage in a scale model sequence. Black points represent all analyzed points to find the white primitives. Colored points are those selected as primitive center.



Fig. 11. Graphic insertion with camera calibration using tripod geometry and primitive tracking.

6 Conclusions

In this paper we study how to improve the primitive tracking stage of the difficult problem of camera calibration of video sequences. These videos are scenarios where, in each frame, there are usually a small number of visible primitives which can be considered to perform the calibration. To track these primitives through the whole video sequence we propose a new method for primitive tracking based on a CART (Classification and Regression Tree). Decision tree is estimated using a learning procedure based RGB image channels and a training set. We present some experiments using HD videos of sport events (soccer matches) in both, scale soccer court models and real scenarios. The procedure we propose is very fast and accurate. Using a combination of RGB channel information the maximum classification error obtained is just about 0'16% (for images which are not included in the training set). As an application of the proposed method, we calibrate the video sequence using the obtained primitive tracking and we illustrate the obtained results showing some images where we have inserted graphics objects in some images using the obtained calibration.

Acknowledgement

This research has partially been supported by the MICINN project reference MTM2010-17615 (Ministerio de Ciencia e Innovación. Spain). We acknowledge MEDIAPRODUCCION S.L. for providing us with the real HD video we use in the numerical experiments.

References

- 1. L. Breiman, JH. Friedman, RA. Olshen, CJ. Stone: Classification and Regression Trees. Belmont, CA: Wadsworth, 1984.
- 2. D. Pea: Anlisis de datos multivariantes. Madrid, 2002.
- 3. T. Hastie, R. Tibshirani, Jerome Friedman: The elements of Statistical Learning. Canada, 2001.
- JB. Hayet and J. Piater: On-Line Rectification of Sport Sequences with Moving Cameras. In: MICAI 2007: Advances in Artificial Intelligence, volume 4827, pages 736-746, 2007.
- H. Kim and KS. Hong: Robust Image Mosaicing of Soccer Videos using Self-Calibration and Line Tracking. In: Pattern analysis and applications, volume 4, pages 9-19, 2001.
- D. Farin and S. Krabbe and PHN. de With and W. Effelsberg: Robust camera calibration for sport videos using court models. In: Storage and Retrieval Methods and Applications for Multimedia, volume 5307, pages 80-91, 2004.
- D. Farin and J.G. Han and PHN. de With: Fast camera calibration for the analysis of sport sequences. In: IEEE International Conference on Multimedia and Expo (ICME), volume 1-2, pages 482-485, 2005.
- Y. Watanabe and M. Haseyama and H. Kitajima: A soccer field tracking method with wire frame model from TV images. In: 2004 International Conference on Image Processing, volume 1-5, pages 1633-1636, 2004.
- M. Aleman-Flores and L. Alvarez and P. Henriquez and L. Mazorra: Morphological Thick Line Center Detection. In: 7th International Conference on Image Analysis and Recognition, volume 6111, pages 71-80, 2010.
- M. Emre Celebi and H. Iyatomi and W. V. Stoecker and R. H. Mossd and H. S. Rabinovitz and G. Argenziano and H. P. Soyer: Automatic detection of blue white veil and related structures in dermoscopy images. In: Computerized Medical Imaging and Graphics, volume 32, pages 670-677, 2008.
- G. Macchiavello, G. Moser, G. Boni, S. B. Serpico: Automatic unsupervised classification of snow-covered areas by decision-tree classification and minimum error thresholding. In: IEEE International Geoscience and Remote Sensing Symposium, volumes 1-5, pages 1251-1254, 2009.
- Z.Zili, Q.Qiming, G. Junping, D. Yuzhi, Y. Yunjun, W. Zhaoqiang, D. Fanwei: CART-Based Rare Habitat Information Extraction For Landsat ETM+. In: IEEE International Geoscience and Remote Sensing Symposium, volume 3, pages 1071-1074, 2008.
- M. Ozuysal, M. Calonder, V. Lepetit, P. Fua: Fast Keypoint Recognition using Random Ferns. In: IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010.
- V. Lepetit, P. Fua: Keypoint Recognition using Randomized Trees. In: IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 28, Nr. 9, pp. 1465-1479, 2006.