Grass Detection for Picture Quality Enhancement of TV Video

DOI:
10.1007/978-3-540-74607-2_62

Document status and date:
Published: 01/01/2007

Document Version:
Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Grass Detection for Picture Quality Enhancement of TV Video

Bahman Zafarifar[1,2,3] and Peter H. N. de With[1,2]

[1] Eindhoven University of Technology, PO Box 513, 5600 MB, The Netherlands
{B.Zafarifar,P.H.N.de.With}@tue.nl
[2] LogicaCMG, PO Box 7089, 5600 JB Eindhoven, The Netherlands
[3] Philips Innovative Applications (CE), Pathoekeweg 11, 8000 Bruges, Belgium

Abstract. Current image enhancement in televisions can be improved if the image is analyzed, objects of interest are segmented, and each segment is processed with specifically optimized algorithms. In this paper we present an algorithm and feature model for segmenting grass areas in video sequences. The system employs adaptive color and position models for creating a coherent grass segmentation map. Compared with previously reported algorithms, our system shows significant improvements in spatial and temporal consistency of the results. This property makes the proposed system suitable for TV video applications.

1 Introduction

Image enhancements in current flat display TVs are performed globally (on the entire image)as in the conventional contrast and brightness adjustments, or locally (on a selected part of the image) as in sharpness enhancement, considering the local statistical properties of the image. For example, some enhancement filters operate along the edge axis, or select a partial set of pixels that are likely to be part of a single object [1]. The local adaptation is typically based on simple pictorial features of the direct neighborhood, rather than considering the true semantic meaning of the object at hand. It is therefore understandable that the obtained picture quality is sub-optimal as compared to a system that locally adapts the processing to the true nature of the objects. Object-based adaptation can be realized if the image is analyzed by a number of object detectors, after which object are segmented and processed with optimized algorithms [2]. Having object detectors in a TV system also enables semantic-level applications such as indoor/outdoor classification, sports detection, semantic-based selection of the received or stored video, or aiding the emerging 3D-TV systems.

Grass fields are frequently seen in TV video, especially in sports programs and outdoor scenes. At the pixel level, grass detection can be used for color shifting and sharpness enhancement, and preventing spurious side effects of other algorithms such as the unintended smoothing effect of noise reduction algorithms in grass areas, by dynamically adapting the settings of the noise filter.

TV applications require that the detection results are pixel-accurate and spatially and temporally consistent, and that the algorithm allows for real-time

J. Blanc-Talon et al. (Eds.): ACIVS 2007, LNCS 4678, pp. 687–698, 2007.

implementation in an embedded environment. Spatial consistency means that the segmentation results should not contain abrupt spatial changes when this is not imposed by the values of the the actual image pixel. Video applications also demand that the segmentation results do not exhibit abrupt changes from frame to frame when the actual image does not contain such abrupt changes. We refer to the latter as temporal consistency. Our algorithm takes these requirements into account and produces a probabilistic grass segmentation map based on modeling the position and the color of grass areas.

The remainder of the paper is organized as follows. In Section 2 we review the previously reported work on real-time grass segmentation for TV applications. Section 3 discusses the properties of grass fields and the requirements of TV applications, Section 4 describes the proposed algorithm, Section 5 presents the results and Section 6 concludes the paper.

2 Related Work

Previously reported work on grass detection for real-time video enhancement includes a method [3] that is based on pixel-level color and texture features. The color feature is in the form of a 3D Gaussian function in the YUV color space, and the texture feature uses the root-mean-square of the luminance component. These two features are combined to form a pixel-based continuous grass-probability function. Due to the pixel-based approach of this method, the resulting segmentation contains significant noise-like local variations, caused by the changing texture characteristics in grass fields. As a result, a post-processed image using this method can contain artifacts due to the mentioned local variations in the segmentation map.

As a solution to this problem, [4] proposes to average the results of a pixel-based color-only grass-detection system using blocks of 8×8 pixels. The obtained average values are then classified to grass/no-grass classes using a noise-dependent binary threshold level. Although the applied averaging alleviates the previously mentioned problem of pixel-level local variations in the segmentation map, the proposed hard segmentation causes a different type of variations in the segmentation result, namely in the form of the nervousness of the resulting 8×8 pixel areas. Such hard segmentation is obviously inadequate for applications like color shifting. Even for less demanding applications like noise reduction, we

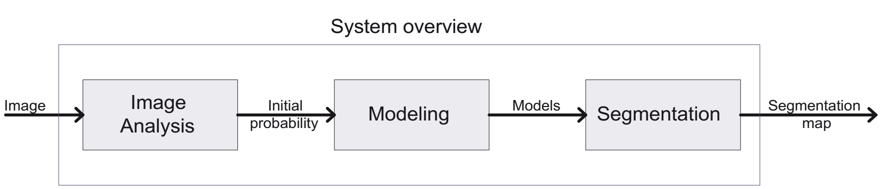

Fig. 1. Overview of the proposed system: starting with image analysis, followed by modeling the color and position of grass areas, and finally segmenting the grass pixels

found that the hard segmentation leads to visible artifacts in the post-processed moving sequences.

We propose a system that builds upon the above-mentioned methods, thereby benefiting from their suitability for real-time implementation, while considerably improving the spatial and temporal consistency of the segmentation results. The proposed system (Fig. 1) performs a multi-scale analysis of the image using color and texture features, and creates models for the color and the position of the grass areas. These models are then used for computing a refined pixel-accurate segmentation map when such accuracy is required by the application.

3 Design Considerations

3.1 Observation of Grass Properties

Grass fields can take a variety of colors, between different frames or even within a frame. The color depends on the type of vegetation, illumination and shadows, patterns left by lawn mowers, camera color settings, and so on. Consequently, attempting to detect grass areas of all appearances is likely to result in a system that erroneously classifies many non-grass objects as grass (false positives). For this reason, we have limited ourselves to green-colored grass (commonly seen in sport videos).

Despite having chosen a certain type of grass, the color can still vary due to shadows. We address this by accounting for color variations within the image, with a spatially-adaptive color model that adapts to the color of an initial estimate of grass areas.

The typical grass texture is given by significant changes in pixel values. The variations are most prominent in the luminance (Y component in YUV color space), and exist far less in the chrominance (U and V) components (see Fig. 3). This high-frequency information in chrominance components is further suppressed by the limited chrominance bandwidth in recording and signal transmission systems [5]. To make the matters worse, the chrominance bandwidth limitation in digitally coded sources often leads to blocking artifacts in the chrominance values of the reconstructed image, resulting in spurious texture when the chrominance components are used for texture analysis. Therefore, we use only the luminance component for texture analysis.

The characteristics of grass texture varies within a frame, based on the distance of the grass field to the camera, camera focus and camera motion. To capture a large variety of grass texture, we employ a multi-scale analysis approach. Grass texture can vary locally due to shadows caused by other grass leaves, or due to a local decrease in the quality of the received signal (blocking artifacts or lack of high frequency components). Therefore, we perform a smoothing operation on the created models to prevent the mentioned local texture variations from abruptly influencing the segmentation result.

3.2 Application Requirements and Implementation Considerations

Our primary target is to use our grass detector for high-end TV applications, such as content-based picture quality improvement. This means that the algorithm should allow for real-time operation, that it should be suitable for implementation on a resource-constrained embedded platform, and that the detection results should be spatially and temporally consistent to avoid artifacts in the post-processed image. We have considered the above-mentioned issues in the design of our algorithm.

- Firstly, we have chosen for filters that produce spatially consistent results and yield smooth transitions in the color and position models.
- Secondly, we have avoided using image-processing techniques that require random access to image data. This allows for implementation of the algorithm in a pixel-synchronous system. The reason behind this choice is that video-processing systems are often constructed as a chain of processing blocks, each block providing the following one with a constant stream of data, rather than having random memory access.
- Thirdly, we have avoided processing techniques that need large frame memories for (temporary) storage of the results. For example, the results of the multi-scale analysis are directly downscaled to a low resolution (16 times lower than input resolution), without having to store intermediate information.
- Lastly, we perform the computationally demanding operations, such as calculations involved in model creation, in the mentioned lower resolution. This significantly decreases the amount of required computations.

4 Algorithm Description

In this section, we describe the proposed system in detail. The system is comprised of three main stages, as shown in Fig. 1. The *Image Analysis* stage computes a first estimate of the grass areas. We call this the *initial probability* of grass. Using this initial probability, we create two smooth models in the *Modeling* stage for the color and the position of the grass areas. While the position model can be directly used for certain applications like adaptive noise reduction or sharpness enhancement, other applications, such as color shifting, require a pixel-accurate soft segmentation map. The *Segmentation* stage calculates this pixel-accurate final segmentation map, using the created color and position models and the image pixel values. The following sections elaborate on the mentioned three stages.

4.1 Image Analysis

In Section 3, we observed that grass areas can take a variety of colors due to illumination differences (shadows, and direct or indirect sunlight). RGB and YUV are the two common color formats in TV systems. In an RGB color system,

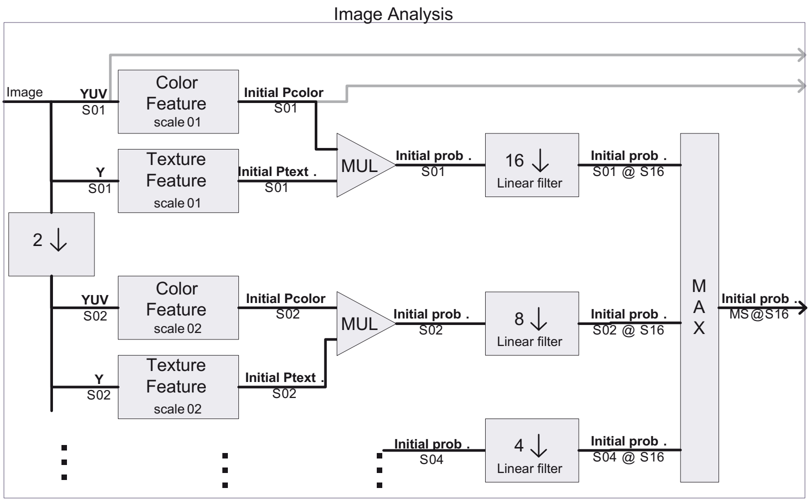

Fig. 2. Schematic overview of image analysis stage. The initial grass probability is calculated for the image in three scales. The results are downscaled and combined to produce the multi-scale initial grass probability.

each component is a function of both chrominance and luminance, while the luminance and chrominance information in a YUV color system are orthogonal to each other. This means that the UV components are less subject to illumination, and therefore we chose the YUV color system for image analysis.

Color: Despite the inherent separation of luminance and chrominance information in the YUV color format, we observed a slight correlation between the luminance and chrominance components for grass areas. Figure 3 depicts the histograms of grass-pixel values in the YUV domain, where the correlation between luminance and chrominance can be seen in the left-most (YU) graph. Our purpose is to approximate this cloud of pixels, using a 3D Gaussian function. This is done by estimating the parameters of this 3D Gaussian using Principle Component Analysis in the training phase. The parameters consist of the center (mean grass color), the orientation of the main axes and the variance along these axes. During the analysis phase, the pixel values, (Y, U, V) are translated by the mentioned mean grass color, and rotated by the axes angles to create the transformed values Y_r, U_r, V_r. The color probability (P_{color}) is then computed by

$$P_{color} = e^{-\left(\left(\frac{Y_r}{\sigma_{y1}}\right)^2 + \left(\frac{U_r}{\sigma_{u1}}\right)^2 + \left(\frac{V_r}{\sigma_{v1}}\right)^2\right)}, \tag{1}$$

where σ_{y1}, σ_{u1} and σ_{v1} are the standard deviations of the corresponding axes.

Texture: Texture is a frequently-used feature in image-segmentation applications [6]. In case of grass detection, the texture feature helps in distinguishing

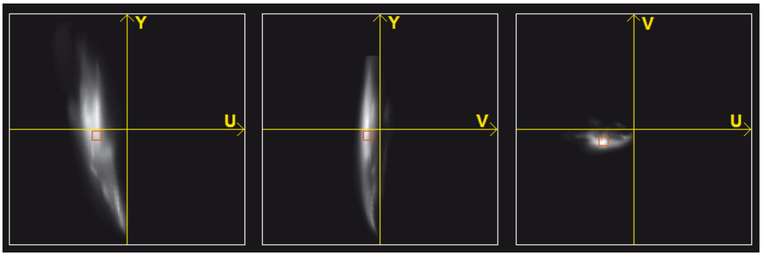

Fig. 3. Histogram of grass-pixel values in the YUV domain, taken over grass areas of a training set, including cloudy, sunny and shadow conditions. Left: U vs. Y, Middle: V vs. Y, Right: : U vs. V.

grass areas from other green objects. In Section 3.1 we motivated the choice of the luminance component for texture analysis. We found that grass has a random, noise-like texture and does not show any unique spatial regularity. In fact, we did not find a way for general distinction between the grass texture and the image noise. Therefore, we subtract the texture measured from image noise from the total measured texture in our texture calculation. As a result, the grass texture can be masked by image noise when the amount of noise exceeds the measured grass texture. For this reason, the texture feature is only useful for images containing a moderate amount of noise. Additionally, the texture feature will provide little information when grass images are taken from a very far distance, or when the quality of the video material is low.

Despite these limitations, texture was found to be a useful feature for separating grass from smooth grass-colored surfaces. As texture measure, we use the Sum of Absolute Differences (SAD) between adjacent pixels in a 5×5 pixels analysis window. The texture metric P_{SAD} is calculated as

$$\begin{aligned} SAD_{hor}(r,c) &= \sum_{i=-w}^{w} \sum_{j=-w}^{w-1} \left| Y(r+i,c+j) - Y(r+i,c+j+1) \right|, \\ SAD_{ver}(r,c) &= \sum_{i=-w}^{w-1} \sum_{j=-w}^{w} \left| Y(r+i,c+j) - Y(r+i+1,c+j) \right|, \\ P_{SAD} &= \frac{SAD_{hor} + SAD_{ver} - T_{SAD}}{N_{SAD}}, \end{aligned} \qquad (2)$$

where SAD_{hor} and SAD_{ver} are the horizontal and vertical SADs respectively, and T_{SAD} is a noise-dependent threshold level. Further, r and c are the coordinates of the pixel under process, w defines the size of the analysis window, and factor $1/N_{SAD}$ normalizes the SAD to the window size. P_{SAD} is further clipped and normalized to a maximum value so that it has the nature of a probability ($P_{texture}$). In the remainder of this paper, we will refer to $P_{texture}$ as a probability.

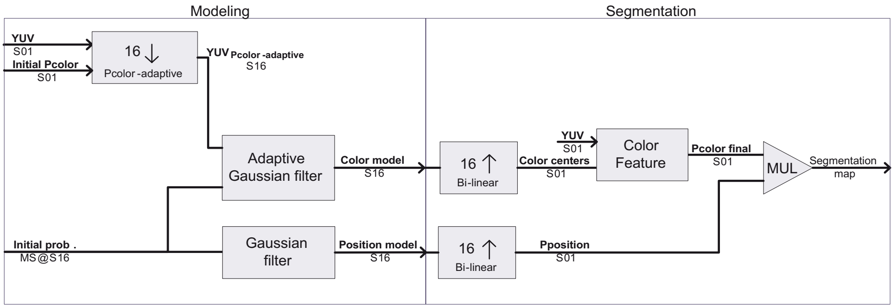

Fig. 4. Modeling and Segmentation stages of the algorithm. Left - Modeling: creating the color and the position models using the initial grass probability. Right - Segmentation: pixel-accurate soft segmentation of grass areas.

Multi-scale Analysis: In Section 3 we observed that the grass texture contains local variations caused by the camera focus, shadows and local image-quality differences (in digitally coded material). In order to capture the grass texture under these different conditions, we have adopted a multi-scale (multi-resolution) image-analysis approach. Using multi-scale analysis, the texture that is not captured in one analysis scale, may still be captured in another scale. Figure 2 depicts the mentioned multi-scale image analysis. Here, the initial grass probability is calculated for three different scales of the image, the image in each scale being half the size of the image in the previous scale. The resulting grass probabilities ($Initialprob.S01, S02, S04$ in Fig. 2) are then downscaled to a common resolution ($Initialprob.S01@S16, S02@S16, S04@S16$ at the right-hand side in Fig. 2) and combined together using the $Maximum$ operation (MAX block in Fig. 2) to produce the multi-scale initial grass probability ($Initialprob.MS@S16$ in Fig. 2). The reason for downscaling is to limit the computation and memory requirements in the modeling stage. The downscale factor (16) was chosen as a tradeoff between lower computation and memory requirements, and spatial resolution of the models, when the input image has Standard Definition resolution.

Three scales of analysis proved to be sufficient for capturing the grass texture. Using lower resolutions for image analysis will lead to a reduced spatial resolution of the initial grass probability, causing spatial inaccuracy of the position- and color models and the eventual segmentation map.

We have considered several measures to reduce the computational complexity and the required memory. Firstly, the calculated initial probabilities of all scales are directly downscaled to a low common resolution ($S16$ in Fig. 2). Secondly, by avoiding the need to store the intermediate (higher resolution) results in the memory, we achieve a high memory efficiency. Thirdly, modeling stage operates on lower resolution images, which considerably decreases the amount of required computations.

For improving the performance of the aforementioned downscaling of the initial probabilities, we use a linear-filtering operation that works as follows. A pixel in the higher-resolution image (the input of the downscaled block) will affect the values of nine pixels of the low-resolution image according to a linear weighting function. The weight is proportional to the the distance between the position of the high-resolution pixel and the centers of the low-resolution pixels. The down-scaled image obtained by this filtering method proved to be much more suitable for moving video material, as compared to block averaging.

4.2 Modeling Grass

Color Model: In Section 3 we noticed that the grass is subject to different illumination conditions. Using fixed color-centers for the final color feature (Fig. 4-right) will lead to partial rejection of grass areas of which the color significantly deviates from the color centers. We found that a better result can be achieved by accounting for the color variation within an image using a spatially-adaptive color model. The model in fact prescribes the expected color of the grass for each image position. To this end, each color component (Y, U, and V) of the image is modeled by a matrix of values of which the dimensions are 16 times smaller than the input image resolution. Each matrix is fitted to the corresponding color component of the image using an adaptively weighted Gaussian filter that takes the initial grass probability as a weight.

The calculation steps are as follows. First, the image is downscaled to the size of the model, using color-adaptive filtering (denoted as $YUV_{Pcolor-adaptive}@S16$ in Fig. 4-left). The color-adaptive filter reduces the influence of outliers, such as extremely bright pixels caused by glair of the sun, on the values of the downscaled image. The downscaled luminance component $Y(r,c)$ is given by

$$Y(r,c) = \frac{\sum_{i=0}^{15}\sum_{j=0}^{15}(Y_{S01}(16r+i,16c+j) \times P_{colorS01}(16r+i,16c+j))}{\sum_{i=0}^{15}\sum_{j=0}^{15}(P_{colorS01}(16r+i,16c+j))}, \quad (3)$$

where Y_{S01} is the luminance component at the input resolution, $P_{colorS01}$ is the color probability at the input resolution, and r and c are the position-indices of the downscaled image.

Next, the color model is computed, using the downscaled representations by (we present only the Y model, M_Y)

$$M_Y(r,c) = \frac{\sum_{i=-h}^{h}\sum_{j=-w}^{w}(Y(r+i,c+j) \times P_{grassInit}(r+i,c+j) \times G(i,j))}{\sum_{i=-h}^{h}\sum_{j=-w}^{w}(P_{grassInit}(r+i,c+j) \times G(i,j))}, \quad (4)$$

where Y is the downscaled luminance component, $P_{grassInit}$ is the initial grass probability, G is a 2D Gaussian kernel, h and w are the model dimensions, and r and c are the model position-indices.

Position Model: We noted in Section 3 that the texture of grass fields contains micro-level variations. Achieving a spatially-consistent detection result requires filtering of these local texture variations. Therefore, we model the positional probability of the grass areas using a smooth position model. The position model $M_{position}$ is obtained by filtering the initial grass probability $P_{grassInit}$ using a Gaussian kernel G as

$$M_{position}(r,c) = \frac{\sum_{i=-l}^{l} \sum_{j=-l}^{l} (P_{grassInit}(r+i, c+j) \times G(i,j))}{\sum_{i=-l}^{l} \sum_{j=-l}^{l} (G(i,j))} , \qquad (5)$$

where l is the size of the Gaussian kernel, and r and c are the model position-indices.

The above-mentioned filtering procedures (Eqns. (3), (4) and (5)) use the computationally demanding division operation. However, the total amount of computations is significantly reduced thanks to the small dimensions of the models (16 times smaller than the input resolution, in both horizontal and vertical dimensions).

Furthermore, to achieve a better temporal stability for moving images, we employ recursive temporal filtering while computing the models.

4.3 Segmentation

When the position model is upscaled to the input image resolution, it produces a map indicating the positional probability of grass for all image positions. This probability map can be directly used for applications like adaptive noise reduction or sharpness enhancement. Other applications, such as color enhancement, may require a pixel-accurate segmentation map, which can be computed as (Fig. 4-right)

$$P_{grassFinal} = P_{colorFinal} \times P_{position} \; . \qquad (6)$$

Here, $P_{position}$ denotes the upscaled version of the position model. $P_{colorFinal}$ is the pixel-accurate final color probability, computed by a 3D Gaussian probability function that uses the YUV values of the image at the input resolution. In contrast to the color feature used in the image analysis-stage (Eqn. (1)), the center of the 3D Gaussian is not fixed here, but defined by the upscaled version of the spatially varying color model. The standard deviations of the 3D Gaussian are smaller than those applied in the image-analysis stage, which helps in reducing false acceptance of non-grass objects. Further, the texture measure has been excluded in the final grass probability to improve the spatial consistency of the detection.

As can be seen in Fig. 4-right, the color and the position models are upscaled (interpolated) by a bi-linear filter prior to being used for determining the color probability. This interpolation is performed on-the-fly, without storing the upscaled images in a memory.

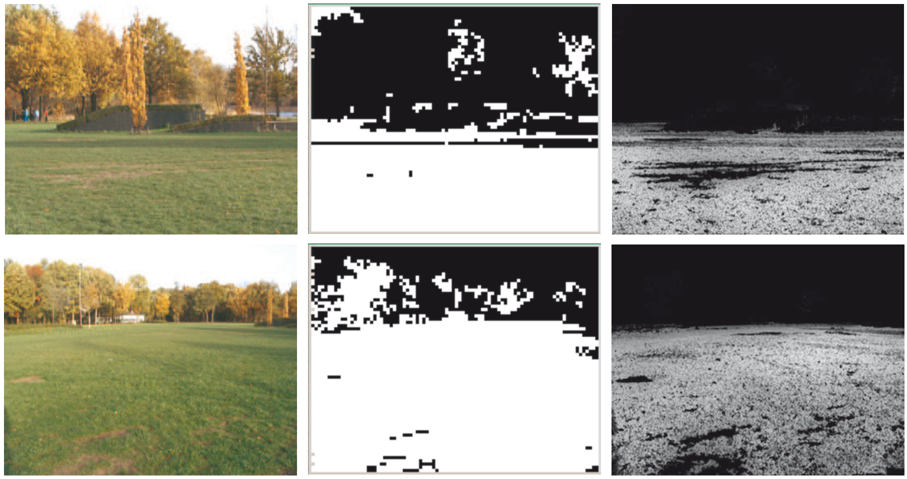

Fig. 5. Results comparison. Left: input, Middle: proposed in [4], Right: our proposal.

5 Experimental Results and Performance Discussion

The proposed algorithm can be trained for detecting grass of a certain color range by choosing appropriate parameters for the color feature. For obtaining these parameters for green-colored grass, we manually annotated the grass areas in 36 images, which were captured under different illumination conditions such as under cloudy and sunny sky, or with and without shadows. Using Principle Component Analysis, we obtained the center, the orientation and the standard deviations of the three axes the 3D Gaussian envelop around the annotated grass pixels (see Fig. 3). We applied the trained algorithm to a test set containing 50 still images and 5 moving sequences, visually inspected the results and made a side-by-side comparison with the algorithm proposed in [4]. The reason for this subjective comparison is that we aim at an algorithm having a high spatial and temporal consistency in the detection result, and at present, there is no metric for such a performance requirement.

Compared with the existing algorithms, we observed a significant improvement in the spatial and temporal consistency of the segmentation results, and improved detection results in images containing grass with different illuminations. We also found the proposed smooth probabilistic segmentation map to be more adequate for image post-processing applications. In the following, we discuss a few examples of the results.

Figure 5 compares the results of our proposal with that of [4]. We can see in the middle column that the existing algorithm detects some tree areas as grass (false positives). Similarly, false positives are found in the ground areas in the middle of the grass field. Our proposal shows a clear improvement in these areas. The improvement is due to a more compact modeling of the grass color values, using the PCA analysis.

Fig. 6. Results comparison. Left: input, Middle: proposed in [4], Right: our proposal.

Fig. 7. Results of the spatially-adaptive color model and the smooth position model. Top-Left: input image, Top-middle: the position model, Top-right: the color model, Bottom-left: segmentation result using fixed color model, Bottom-middle: segmentation result using spatially adaptive color model, Bottom-right: result existing algorithm.

Figure 6 portrays a more complex, which is difficult for both algorithms. First, we notice the false positives of the existing algorithm in the flower garden, whereas these small green objects are filtered out in our proposal owing to the smooth position model. Second, we notice that both algorithms have problems with the tree areas at the top of the picture. Such false positives occur in our algorithm on large, green textured areas (tree leaves). Lastly, we notice that our algorithm produces lower probabilities in the smooth grass area at the top-right side of the image, resulting in missing grass detection in that area. This is due to the absence of texture in these areas. This false negative is not in the form of abrupt changes, making the consequences less severe.

Figure 7 shows the benefit the adopted locally adaptive color model. We can see that although there is a large difference in the color of sunny and shadow areas, the resulting segmentation map (Bottom-middle) does not abruptly reject any of these two areas. While the existing algorithm (Bottom-right) shows a deteriorated detection in the shadow, our algorithm (Bottom-middle) preserves a positive detection of grass, albeit at a lower probability.

6 Conclusion

We have presented an algorithm for consistent detection of grass areas for TV applications, with the aim to improve the quality in the grass areas in the image. For such applications, it is of utmost importance that the image segmentation results are both spatially and temporally coherent. Not complying with this requirement would lead to artifacts in the post-processed video. To achieve this, we have modeled the grass areas using a spatially adaptive color model and a smooth position model. The color model accounts for the large color range of the grass areas within the image, which occurs particularly when the image contains both sunny and shadowed parts. The position model ensures that local variations of the grass texture do not abruptly influence the segmentation result. Furthermore, a multi-scale image analysis approach helps in capturing different appearances of grass. When compared to an existing algorithm, our system shows significant improvements in spatial and temporal consistency of the segmentation result.

During the algorithm design, we kept the limitations of an embedded TV platform into account. As such, we avoid the need for storing intermediate results by directly downscaling the analysis results to a low resolution, and by performing the more complex computations at this low resolution. This approach decreases the memory and computation requirements. Furthermore, the algorithm is suitable for implementation in a pixel-synchronous video platform. This is due to our choice for analysis and modeling techniques which have a regular memory access and deterministic computation requirement, as compared to techniques that require random access to image data, or exhibit a variable computation demand.

Acknowledgement

The authors gratefully acknowledge Dr. Erwin Bellers and Stephen Herman for their specific input on the existing algorithms for real-time grass detection.

References

1. de Haan, G.: Video Processing for Multimedia Systems. University Press, Eindhoven (2000)
2. Herman, S., Janssen, J.: System and method for performing segmentation-based enhancements of a video image, European Patent EP 1 374 563, date of publication (January 2004)
3. Herman, S., Janssen, J.: Automatic segmentation-based grass detection for real-time video, European Patent EP 1 374 170, date of publication (January 2004)
4. Herman, S., Bellers, E.: Image segmentation based on block averaging, United States Patent US 2006/0072842 A1, date of publication (April 2006)
5. Netravali, A., Haskell, B., Puri, A.: Digital Video: an Introduction to MPEG-2. International Thompson Publishing (1997)
6. Alan, C.: Handbook of Image and Video Processing. Academic Press, London (2000)