Multiresolution Color Image Segmentation Applied to Background Extraction in Outdoor Images

Sébastien Lefèvre [1,2], Loïc Mercier [1], Vincent Tiberghien [1], Nicole Vincent [1]
[1] Laboratoire d'Informatique, E3i / Université de Tours,
64, avenue Portalis,
37200 Tours, France
[2] AtosOrigin,
19, rue de la Vallée Maillard,
BP1311, 41013 Blois Cedex, France

Abstract

An adaptive technique for color image segmentation is presented in this paper. The segmentation is performed using a multiresolution scheme and considering the background areas have quite uniform color features at a low-resolution representation of the image. First, a pyramidal representation of the original image is built. Then segmentation is improved iteratively at each resolution using color information. This method allows to extract background areas in outdoor images. Background and foreground separation is especially useful in initialization of object tracking applications. Several color spaces are compared in order to determine a robust method specially with respect to illumination changes which frequently occur in outdoor images. Mean value of Hue component from HSV (Hue, Saturation, Value) color space is selected as the best decision criterion for image matching. Segmentation results of color image from soccer game video sequences are presented to illustrate the method efficiency.

1. Introduction

In most images, either still images or video images, the aim is to show something or someone that happens to be present within an environment. The purpose is not to show the surrounding but rather the point of interest, the foreground. Besides, to have a well balanced image it is often better to have neither a too important foreground nor too small points of interest. The background is also chosen such that it can easily be differentiated from the core of the image. We are here working on such images. That is to say we are not interested in face images where the face fills up the whole image. We are more concerned with images present in video sequences were some moving objects are looked at. The observation field has to be large enough in order to handle at any time the pieces of interest.

Among video understanding applications, object tracking has quite a major place. This can be seen as a segmentation task between background and foreground, or as an extraction task of the background. The learning of the characteristics of the background areas is not always an easy step. In fact it can evolve along the video sequence. The same is true for the foreground areas that may be unknown. The goal of this processing is to separate background and foreground areas in images. In case of an object tracking application, foreground areas can then be used for object initialization.

Several techniques have been proposed for background extraction. In case of a static camera, most common methods are based either on successive frame comparison (computing absolute difference) or comparison with a reference frame. In the last case, we assume no object is present in the scene when recording the reference frame, which can also be updated through time in order to take into account illumination modifications. Depending on the application some differences within the background may occur and must not induce the detection of a foreground element as far as only limited fuzzy modifications are concerned.

When dealing with moving camera, a motion compensation step has to be first performed [6] but these methods are often characterized by a high computation cost. Then the problem can be solved using similar methods as in the case of static camera. Several authors have proposed to include additional information, such as range images [2] that introduce depth information or stereo images [3] that enable 3D reconstruction. Some real time systems have also been proposed [5]. Evolution along time brings an important information, nevertheless, image by image information is some time necessary before hand in order to initiate a process.

In case of still image analysis, background model can not be improved through time and no motion information is available. The method we propose in this paper deals with color image to perform background extraction. But as we intend to reach real time processing of video sequences, the computational cost has to be minimized. Then it is of interest to work with a single color image. We have made choice of a 1-Dimensional feature to express the color image. Then, due to its low computational cost, it can be used in an object tracking application as a faster alternative to classical background extraction method based on motion compensation followed by image difference.

Using a multiresolution scheme, a background model is first obtained from low-resolution image in a adaptive way. Then the segmentation is improved iteratively based on an image matching process. In order to deal with natural (outdoor) images, a comparative study of several color spaces is necessary to select a decision criterion robust to illumination change. In fact we are particularly interested in outdoor images where the illumination changes are more obvious than in indoor sceneries. Indeed the problem of light is present in almost any case and we try to minimize the influence on the segmentation results we present.

In a first part we will see how a multiresolution approach can be applied that allows a learning phase of the characteristics. Next we will discuss about a decision criterion that makes possible the labeling of subimages through a matching process. A comparative study of several color spaces will be developed. Finally results of the proposed segmentation method applied on soccer game images are presented.

2. Multiresolution Image Segmentation

Among the so many segmentation methods we have privileged a multiresolution method. It makes possible to adapt the preciseness of the segmentation result to the need of the following processing in the application. In the general case it has the advantage to be sparse in terms of computational resources that is an important point when video sequences are concerned. The method we propose allows to divide an image into several areas depending on different criteria, following a multi-resolution scheme with a learning step and a coarse to fine approach.

First the image is analyzed at a low (or coarse) resolution to obtain a preliminary result. This result is then iteratively improved using higher (or finer) resolution. A multiresolution representation can be obtained using many ways. For example a wavelet transform can be used [4,7,9], but it is a rather time consuming approach we have not chosen. We have preferred to develop a pyramidal decomposition [1]. This kind of approach is often characterized by computational cost lower than for global (only one resolution level) segmentation methods. The main problems linked with these multiresolution approaches come from the determination of the right levels to begin with or to stop the study.

The first step in the proposed method consists in creating the multiresolution representation of the image. This is performed using a pyramidal model where every pixel at resolution $n + 1$ is computed as the average of a set of s pixels at resolution n. From an original image whose resolution is referred by a zero value of n, we finally obtain a low resolution image with n equals to n_{max}.

Once the pyramidal representation of the image has been computed, it is possible to determine the background model used in the segmentation process. We consider the background is modeled by the lowest resolution image ($n = n_{max}$). The assumption comes from the fact that when we look at an image from a far viewpoint we mainly see the image background. When the point of view becomes closer, we not only see background but also foreground objects. Of course this only holds when the part occupied by the background is significant in the image and the foreground objects are different enough from the background. The non global uniformity of the background could be handled using a clustering process at low resolution with a morphological supervision of the cluster shapes.

The segmentation can then be performed and improved iteratively from resolution $n_{max} - 1$ to initial resolution ($n = 0$). Let us suppose the process has been applied from level n_{max} where the background model has been built to level k using $n_{max} - k$ iterations. Then each element of the k-level image has been labeled as foreground and background. The model of the background can be improved by averaging its elements. The segmentation can be propagated at image level k–1. The k-level image is equivalent to the initial image that would be divided into $r^{n_{max}-k}$ regions where r is a constant. Each of these regions are then compared to the background model using a criterion we will precise in the next section. If the matching between a region and the background model is correct, we label this part of the image as background area. Of course at each level only regions which were not labeled as background are processed.

At any level the process is stopped, and regions with no label (which are considered as foreground areas) are analyzed. In case of applications needing very accurate segmentation, foreground areas can be further analyzed in order to improve the object contours. On the opposite, when coarse segmentation is enough for the considered task (*e.g.* object tracking), the process can be stopped at resolution n_{final} with $n_{max} > n_{final} \geq 0$.

Here we have presented the general scheme of the method. Figures 1 and 2 respectively show the pyramidal representation used in the proposed approach and the successive steps of our method. Of course the results are largely depending both on the way the background is modeled and on the criterion used to decide whether a region is "background" labeled or not. So, now let us come to the criterion used in the comparison process of any area and the background.

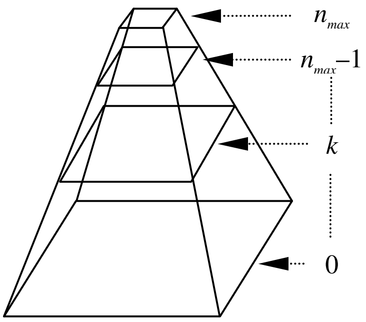


Figure 1: Pyramidal representation: levels 0 and n_{max} represent respectively original resolution and resolution used in background modeling.

- Create the pyramidal representation
- Estimate the background model $C_{mean}(I_{bg})$ (with level equal to n_{max})
- Set level ← level – 1
- While level > n_{final}
  - Create new regions from unlabeled regions at resolution equal to level +1
  - For each region I_r do
    - Compute $C_{mean}(I_r)$
    - If $d(C_{mean}(I_{bg}),C_{mean}(I_r)) < T$
      - Label region I_r as background
  - Set level ← level – 1
- Label unlabeled regions as foreground

Figure 2: Description of proposed multiresolution method (without improving background model iteratively).

3. Decision criterion

In order to compare and match a region of the image at resolution level n with the background model, we have to determine a decision criterion. Our purpose is to obtain a criterion robust with respect to illumination changes. Besides, as we are dealing with color images, we have chosen to use a criterion linked to color. A comparative study of several color spaces has been achieved to define this criterion and is presented in next section. Contrary to Vandenbroucke et al which propose in [8] to use an hybrid color space by selecting most discriminating color component in a given image, we use only one color component to perform image segmentation in order to limit the computation time.

With the same goal and to define the background model, we choose to average the selected color component values of pixels belonging to a region or to the background model. It allows us to keep only one value C_{mean} per region. Some other statistical measures could also be used, but most of the time their processing needs more computational resources than a classical mean measure.

The background model is limited to only one value. Then each region to be compared to the background has to be represented in the same way. That is to say with any region is associated the mean value of the color attribute of the image elements at the level being studied.

The matching condition between a region and the background model involves the use of a threshold T and can then be modeled by the following equation:

$$d(C_{mean}(I_{bg}),C_{mean}(I_r))<T \qquad (1)$$

where d is some distance and $C_{mean}(I_{bg})$ and $C_{mean}(I_r)$ represent respectively the mean values of the background model and of the region considered.

Choice of the color component C will be explained and justified in the following section.

4. Comparative study of color spaces

In order to determine the decision criterion used in the matching process belonging to the proposed multiresolution segmentation method, we perform a comparative study of most commonly used color spaces. Color spaces involved in this study were RGB, normalized RGB (where r + g + b = 1), CMY, XYZ, YUV, HSV, and HSI spaces.

When selecting the color component to be used in the matching process, we have to take into account that we are dealing with background extraction in outdoor images or images that may vary along time as illumination is not constant. In this kind of images, illumination changes can frequently appear. So we have to choose a color component robust to lighting conditions. This selection can be made from theoretical justifications and practical experiments can confirm the choice.

From a theoretical point of view, some color components are known to be insensitive to changing lighting conditions, whereas some others lack of robustness. Among them, graylevel or luminance components (Y in YUV, V or I in HSV and HSI respectively) are directly linked with lighting conditions and so are very sensitive to illumination changes. They have to be avoided. Some other color components involve more or less luminance feature (*e.g.* components of RGB and complementary CMY spaces) and cannot be selected as color component independent of lighting condition.

Because we are using the selected color component as a decision criterion in a multiresolution image segmentation, we have to choose a measure robust to multiresolution processing. More precisely, successive average measures have to be computed. Some color components are less robust than others to this kind of artifact. For example, S component from HSV will be characterized by a shorter range of values as the resolution changes and the averaging effect is higher. This is not the case with H component from the same HSV color space.

Practical experiments were also performed in order to determine the color component used in the matching process. The proposed segmentation approach was performed on several images using every color component of the set of color spaces described previously. Comparison of experimental results with theoretical segmentation led us to use the H (Hue) component from HSV color space in the decision criterion computation. As it was described in theoretical aspect presentation, the hue color component H is robust to illumination change (contrary to the value component V) and also to the successive averaging phase processed in pyramid creation (contrary to the saturation component S).

Selection of Hue value from HSV color space as the color component used in our matching process allows us to precise the matching condition proposed in equation (1), and especially the *d* function:

$$d(H_{mean}(I_{bg}),H_{mean}(I_r))<T \qquad (2)$$

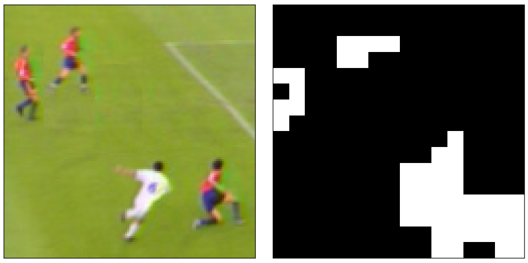
where $H_{mean}(I_{bg})$ and $H_{mean}(I_r)$ represent respectively the mean Hue values of the background model and of the region considered. Hue values are computed as angles (in degrees) belonging to interval [0,360]. The function *d* used in the matching process can then be defined as:

$$d(a,b)=min(abs(a-b),360-abs(a-b)) \qquad (3)$$

Hue color component has been selected as the decision criterion used in the matching process. Theoretical justification and practical experiments lead us to this choice. We will now see results of the proposed multiresolution segmentation method.

5. Results

The method presented here has been tested on outdoor images, where illumination is not constant. Figure 3b shows the segmentation result of a color image extracted from a soccer game video sequence and presented in figure 3a.



(a) *(b)*

Figure 3: Original image (a) and resulting segmentation (b) (black pixels for background, white pixels for foreground)

The size of input image is 128x128 pixels. Grass field has been labeled as background (black pixels) whereas the foreground area (white pixels) contains mainly pixels belonging to soccer players. This segmentation was obtained using parameters presented in Table 1.

Parameters	Description	Values
n	Number of layers in the pyramid	7
n_{final}	Resolution used to obtain final result	5
s	Number of pixels used at resolution n to generate a pixel a resolution $n+1$	4
r	Number of regions	4
T	Threshold for region and background model matching	2

Table 1: Parameters used in the segmentation process

Choice of the final resolution n_{final} or the number of layers n in the pyramid considered in the proposed multiresolution approach has a direct influence in the precision of the resulting image. Figure 4 shows segmentation results obtained after different numbers of iterations are performed.

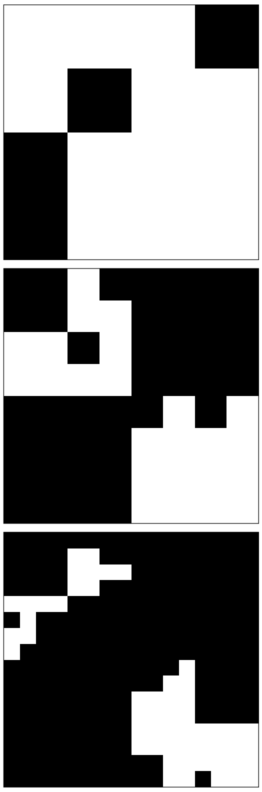


Figure 4: Extracted background and foreground for different final resolution values (n_{final} is equal to 3 to 5 from top to bottom)

6. Conclusion

An adaptive multiresolution image segmentation technique using color information has been presented in this paper. The proposed approach uses Hue component of the HSV color space and by this way is robust to illumination change. This method is dedicated to foreground / background separation in outdoor images and has been successfully applied to soccer game image segmentation. Due to its low computation cost, it can be used as a preprocessing step in an object tracking application. Indeed the segmentation result gives information about initial object positions.

The method presented in this paper deals with images composed of background areas with similar color features. It has to be extended to images with non-uniform background, which is often the case when images include the sky. We have already tested different color spaces, some other can be discussed and future work will also concern testing other comparison criteria. Finally, texture information may also be introduced to improve the segmentation.

References

1. M.C. Comer and E.J. Delp, “Multiresolution Image Segmentation”, *IEEE International Conference on Acoustics, Speech, and Signal Processing*, May 1995, Detroit, MI, USA, pp. 2415-2418.
2. G. Gordon, T. Darell, M. Harville, and J. Woodfill, “Background Estimation and Removal Based on Range and Color”, *IEEE International Conference on Computer Vision and Pattern Recognition*, June 1999, Fort Collins, CO, USA, Vol. 2, pp. 459-464.
3. Y. Ivanov, A. Bobick, and J. Liu, “Fast Lighting Independent Background Subtraction”, *International Journal of Computer Vision*, Vol. 37, No. 2, June 2000, pp. 199-207.
4. S.K. Kopparapu, P. Mudalige, and P. Corke, “A Multiresolution Based Image Segmentation”, IEE International Conference on Image Processing and Its Applications, Manchester, UK, July 1999, pp. 567-571.
5. F.C.M. Martins, B.R. Nickerson, V. Bostrom, and R. Hazra, “Implementation of a Real-Time Foreground / Background Segmentation System on the Intel Architecture”, Workshop on Frame Rate Applications, Methods and Experiences with Regularly Available Technology and Equipment, September 1999, Kerkyra, Greece.
6. R. Mech and M. Wollborn, “A Noise Robust Method for 2D Shape Estimation of Moving Objects in Video Sequences Considering a Moving Camera”, *Signal Processing*, Vol. 66, No. 2, April 1998, pp. 203-217.
7. M.G. Ramos, S.S. Hemami, and M.A. Tamburro, “Psychovisually-Based Multiresolution Image Segmentation”, *IEEE International Conference on Image Processing*, October 1997, Santa Barbara, CA, USA, Vol. 3, pp. 66-69.
8. N. Vandenbroucke, L. Macaire, and J.G. Postaire, “Color Pixels Classification in an Hybrid Color Space”, *IEEE International Conference on Image Processing,* October 1998, Chicago, IL, USA, Vol. 1, pp. 176-180.
9. J.Z. Wang, J. Li, R.M. Gray, and G. Wiederhold, “Unsupervised Multiresolution Segmentation for Images with Low Depth of Field”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 23, No. 1, January 2001, pp. 85-90.

Biography

Sébastien LEFEVRE received the M.S. degree in Computer Science from University of Technology of Compiègne, France, in 1999. He is currently a PhD student in Computer Science at University François Rabelais, Tours, France. He is also in the R&D team of AtosOrigin Customer Management Services, Blois, France. His research interests include real time video indexing, color image / video analysis and processing, and event detection for video understanding.