ANN PATTERN RECOGNITION TECHNIQUES AND VIRTUAL ENVIRONMENTS-BASED FOR IMAGE PROCESSING IN MOBILE ROBOT NAVIGATION

LUCIANO C. LULIO¹, CLAITEM DE OLIVEIRA¹, MARIO L. TRONCO¹, ARTHUR J. V. PORTO¹,

1. Escola de Engenharia de São Carlos / Universidade de São Paulo - EESC/USP (Engineering School of São Carlos, University of São Paulo)

Avenue Travallor São-Carlense, 400, CEP 13566-590, São Carlos, SP - Brazil
E-mails: lclulio@sc.usp.br, claiton@sc.usp.br, mltronco@sc.usp.br, ajvporto@sc.usp.br

Abstract— The main focus in this research is to apply image processing techniques in computer vision to agricultural mobile robots (AMR) used for guidance navigation problems, as well as trajectory matters. Several segmentation algorithms proposed in literature (CHEN et al., 2002; HILL, 2002; LIU, ZHOU, 2004) were designed to process images originally characterized by the above-mentioned items. Therefore, agricultural automation uses computer vision resources, which can be applied to a number of different tasks, such as inspection (BROSNAN and SUN, 2002), classification of plants (TANG, TIAN, STEWARD, 2003; NETO et al., 2003; STEWARD et al., 2004), estimated production (ANNAMALAI, LEE, BURKS, 2004), automated collection (PLEBE, GRASSO, 2001) and guidance of autonomous machines.

As mentioned above and considering such issues, the present project is applied on both JSEG segmentation algorithm (DENG et al., 1999a) and multilayer perceptron (MLP) in order to segment and therefore classify images into the following classes: navigable area, planting area, and sky; and provide a real object data model for automatic generation of virtual environments from natural scenes of agricultural crops.

Keywords— Pattern recognition, Image segmentation, Mobile robots, Computer vision, Nature scenes.

1. INTRODUCTION

Algorithms for crops inspection and analysis, such for planting area avoidance navigation, work mainly with images composed of complex objects, textures, shadows and brightness - as part of unstructured environments. Several segmentation algorithms proposed in literature (CHEN et al., 2002; HILL, 2002; LIU, ZHOU, 2004) were designed to process images originally characterized by the above-mentioned items. Therefore, agricultural automation uses computer vision resources, which can be applied to a number of different tasks, such as inspection (BROSNAN and SUN, 2002), classification of plants (TANG, TIAN, STEWARD, 2003; NETO et al., 2003; STEWARD et al., 2004), estimated production (ANNAMALAI, LEE, BURKS, 2004), automated collection (PLEBE, GRASSO, 2001) and guidance of autonomous machines.

As mentioned above and considering such issues, the present project is applied on both JSEG segmentation algorithm (DENG et al., 1999a) and multilayer perceptron (MLP) in order to segment and therefore classify images into the following classes: navigable area, planting area, and sky; and provide a real object data model for automatic generation of virtual environments from natural scenes of agricultural crops.

In order to achieve such objective, the real object data model was built based on information obtained by a computer vision system. Moreover, the approach tried to segment classification deploys an artificial neural network (ANN) – iRPROP algorithm, faster than usual Back-propagation algorithm (IGEL, HÜSKEL, 2003) – to classify and characterize the segments into three classes – turning real data models an alternative solution for improving the classification of planting lines. A feature vector was formed with color channels histograms. After training, the mean squared error (MSE), which was obtained in the different ANN topologies, indicated the most appropriate topology. Ultimately, the results achieved by segment classification were used to create the image-class map. After segmentation, it was necessary to identify and classify the segments.

2. IMAGE SEGMENTATION STEP 1

Colored images with homogeneous regions are segmented with an unsupervised algorithm to generate clusters in the color space/class (MOREIRA, COSTA, 1996; TIAN, SLAUGHTER, 1998). The way how to segment images with textures is to consider the spatial arrangement of pixels using a region-growing technique whereby a homogeneity mode is defined with pixels grouped in the segmented region. Moreover, in order to segment texture images one must consider different scales of images.
The algorithm named JSEG segments images of natural scenes properly, without manual parameter adjustment for each image and simplifies texture and color. Segmentation with this algorithm passes through three stages, namely color space quantization (number reduction process of distinct colors in a given image), hit rate regions and similar color regions merging.

First of all, the color space is quantized with little perceptual degradation by using the quantization algorithm (DENG, MANJUNATH, SHIN, 1999b) with minimum coloring. Each color is associated with a class. The original image pixels are replaced by classes to form the class maps in the next stage. Before performing the hit rate regions, the J-image - a class map for each windowed color region, whose positive and negative values represent the edges and textures of the processing image - must be created with pixel values used as a similarity algorithm for the hit rate region. These values are called 'J-values' and are calculated from a window placed on the quantized image, where the J-value belongs.

In order to calculate the J-value, Z is defined as the set of all points of quantized image, then \( z = (x, y) \) with \( z \in Z \) and being \( m \) the average in all \( Z \) elements. \( C \) is the number of classes obtained in the quantization. Then \( Z \) is classified into \( C \) classes, \( Z_i \) are the elements of \( Z \) belonging to class \( i \), where \( i=1,...,C \), and \( m_i \) are the element averages in \( Z_i \). The J-value is as follows:

\[
J = \frac{S_T}{S_W} = \frac{(S_T - S_W)}{S_W}
\]

(1)

\[
S_T = \sum_{z \in Z} \| z - m \|^2
\]

(2)

\[
S_W = \sum_{i=1}^{C} \sum_{z \in Z_i} \| z - m_i \|^2
\]

(3)

Parameter \( S_T \) denotes the sum of quantized points, within all \( Z \) elements average. The \( S_W \) relation denotes the distance metrics for random non-linear distribution on class-maps.

Figures 4, 5, 6 and 7 illustrate not only color quantization and spatial distributions of J-image in others natural scenes, but the flood fill implemented algorithm, for determining the boundaries edges connected on the region growing areas (using queue data structure provided from region valleys). All scenes were submitted to a gradient magnitude, as segmentation function rating (Sobel masks for higher values at the borders of navigation areas and lower values inside planting areas), then image is segmented with a watershed transform directly on the gradient magnitude. JSEG outperforms the evaluation for all images, with an effective spatial distribution on planting lines.

Regions with the lowest values of J-image are called valleys. Thus, it is possible to determine the starting point of efficient growth, which depends on the addition of similar valleys. The algorithm ends when there are spare pixels to be added to those regions.
3. SEGMENTS CLASSIFICATION AND ANN

Due to the nature of nonlinear vectors, it is fundamental that an ANN-based pattern classification and recognition be used. Multi-Layer Perceptron (MLP) (HAYKIN, 2001) was implemented through a customized back-propagation algorithm for complex patterns.

Derived from back-propagation, the iRPROP algorithm (improved resilient back-propagation) (CAVANI, 2004) is both fast and accurate, with easy parameter adjustment. It features an Octave (COSTA, CESAR JUNIOR, 2001) module which was adopted for the purposes of this work and it is classified with HSV (H – hue, S – saturation, V – value) color space channels histograms of 256 categories (32, 64, 128 and 256 neurons in a hidden layer training for each color space channel: H, HS, and HSV - denoted by \( m \)). The output layer has three neurons, each of them having a predetermined class (Fig. 8).

![Fig. 3. JSEG image segmentation process 2.](image)

![Fig. 4. Watershed transform of gradient magnitude in flood fill class-map for scene 2.](image)

![Fig. 5. ANN schematic topology.](image)

<table>
<thead>
<tr>
<th>MSE</th>
<th>Neurons (( m ))</th>
<th>Navigation area</th>
<th>Planting area</th>
<th>Sky</th>
</tr>
</thead>
<tbody>
<tr>
<td>HS</td>
<td>32</td>
<td>0.089143</td>
<td>0.094905</td>
<td>0.023409</td>
</tr>
<tr>
<td></td>
<td>64</td>
<td>0.099398</td>
<td>0.045956</td>
<td>0.089776</td>
</tr>
<tr>
<td></td>
<td>128</td>
<td>0.049100</td>
<td>0.095964</td>
<td>0.097455</td>
</tr>
<tr>
<td></td>
<td>256</td>
<td>0.057136</td>
<td>0.099843</td>
<td>0.034532</td>
</tr>
<tr>
<td>HSV</td>
<td>32</td>
<td>0.089450</td>
<td>0.02453</td>
<td>0.067545</td>
</tr>
<tr>
<td></td>
<td>64</td>
<td>0.059981</td>
<td>0.010384</td>
<td>0.082364</td>
</tr>
<tr>
<td></td>
<td>128</td>
<td>0.049677</td>
<td>0.078453</td>
<td>0.043493</td>
</tr>
<tr>
<td></td>
<td>256</td>
<td>0.038817</td>
<td>0.079856</td>
<td>0.045643</td>
</tr>
</tbody>
</table>
All ANN-based topologies were trained with a threshold lower than 0.0001 mean squared error (MSE) (HAYKIN, 1999; COMANICIU, MEER, 1997), the synaptic neurons weights are initiated with random values and the other algorithm parameters were set with Fast Artificial Neural Network (FANN) library (NISSEN, 2006) for Matlab (Mathworks Inc.) platform. The most appropriate segment and topology classifications are those using vectors extracted from HSV color space. Also, a network with less MSE in the HSV-64 was used so as to classify the planting area; for class navigable area (soil), HSV-256 was chosen; as for the class sky, the HS-32. Tab. I shows the previously data.

4. IMAGE SEGMENTATION STEP 2

Among the most common characteristics, such as color, texture, size, shape and position, the proposed methodology uses only the shape, size and position as parameters for the automatic generation of virtual environments (ANNAMALAI, LEE, BURKS, 2004; SILVA, 2008).

The process of extracting the desired characteristics is performed by drawing some specific lines onto the image and then calculating the size of these lines to estimate the size of the object as well as its shape.

As the resulting image from segmentation is binary, some lines of interest are drawn in the original image acquired (RGB image), such as: the contour line, a vertical line passing through the centroid of the object and some horizontal lines perpendicular the vertical line of the object.

4.1 Lines of interest drawing

To draw the contour line of the object it is used the coordinates x and y of the vector which stores the points of the contour. Then, the centroid of the object is determined through the geometrical moments of the contour.

Once the centroid is obtained, a vertical straight line segment must be drawn passing through the centroid of the object of interest dividing it into two hemispheres (left and right), as follows:

Be \( x_c \) and \( y_c \) of the centroid coordinates and \( X, Y \) the contour coordinates:

- the upper end of the vertical line segment is determined by \( (x_c, \min(Y)) \), where \( \min(Y) \) is the smallest value of \( y \) belonging to the contour.
- the lower end of the vertical line segment is determined by \( (x_c, \max(Y)) \), where \( \max(Y) \) is the largest value of \( y \) belonging to the contour.

For each hemisphere (left and right) of the contour, \( n \) segments of straight horizontal lines must be drawn perpendicular to the vertical straight line segment, as follows:

- the distance between the segments of horizontal lines must be constant for all rows. The choice of distance value depends on the level of detail of the silhouette one wants to extract to build the real data model to be used in the correspondent virtual environment. The default value adopted is 10 pixels.
- the number (\( n \)) of segments of horizontal lines must be determined by the ratio between the height of the object (in pixels) and the distance between the horizontal segments to be drawn. The height of the object can be obtained through the distance between the ends of the segment of vertical line that divides the contour into two hemispheres.
- starting from the top of the vertical line segment, for each hemisphere, it must be drawn a horizontal line segment that begins at the vertical line segment to the farthest point of contour found in the image.

The role of the segments of horizontal lines is the same as dividing the actual segmented object into multiple slices to facilitate the generation of the corresponding virtual object based on information obtained from these slices. As the segments of horizontal lines within the limits of the contour, they allow to obtain both the size and the shape of the object of interest, as shown in Figure 6.
Because the information about the objects extracted from natural scenes is formed by numerous and repeated fields and structures, the real object data model (RODM) must have a simple, readable and extensible format. In this context, the XML standard (eXtensible Markup Language) is shown as a suitable format. Therefore, the generation of the RODM consists of the generation of an XML document containing the information of real objects obtained from the computer vision system.

To generate the RODM he XML elements must be organized in a hierarchical manner that facilitates their use during the virtual world generation. The hierarchy of the elements of the XML document generated during the stages of the computer vision system is shown in Figure 7.

The results presented here were obtained from images of natural scenes of orange trees using a CCD video camera in the RGB color space, with resolution of 1920x1080 (high definition) at 24 fps (frames per second).

After the performance of border detection and morphological transformations previously described, the results of outer contours extraction for each object of interest in the natural scene are shown in Figure 8. For each object segmented from the images, some lines of interest were drawn as described previously (Figure 9).

5. REAL OBJECT DATA MODEL

Because the information about the objects extracted from natural scenes is formed by numerous and repeated fields and structures, the real object data model (RODM) must have a simple, readable and extensible format. In this context, the XML standard (eXtensible Markup Language) is shown as a suitable format. Therefore, the generation of the RODM consists of the generation of an XML document containing the information of real objects obtained from the computer vision system.

To generate the RODM he XML elements must be organized in a hierarchical manner that facilitates their use during the virtual world generation. The hierarchy of the elements of the XML document generated during the stages of the computer vision system is shown in Figure 7.

The results presented here were obtained from images of natural scenes of orange trees using a CCD video camera in the RGB color space, with resolution of 1920x1080 (high definition) at 24 fps (frames per second).

After the performance of border detection and morphological transformations previously described, the results of outer contours extraction for each object of interest in the natural scene are shown in Figure 8. For each object segmented from the images, some lines of interest were drawn as described previously (Figure 9).

5. REAL OBJECT DATA MODEL

Because the information about the objects extracted from natural scenes is formed by numerous and repeated fields and structures, the real object data model (RODM) must have a simple, readable and extensible format. In this context, the XML standard (eXtensible Markup Language) is shown as a suitable format. Therefore, the generation of the RODM consists of the generation of an XML document containing the information of real objects obtained from the computer vision system.

To generate the RODM he XML elements must be organized in a hierarchical manner that facilitates their use during the virtual world generation. The hierarchy of the elements of the XML document generated during the stages of the computer vision system is shown in Figure 7.

The results presented here were obtained from images of natural scenes of orange trees using a CCD video camera in the RGB color space, with resolution of 1920x1080 (high definition) at 24 fps (frames per second).

After the performance of border detection and morphological transformations previously described, the results of outer contours extraction for each object of interest in the natural scene are shown in Figure 8. For each object segmented from the images, some lines of interest were drawn as described previously (Figure 9).
6. CONCLUSIONS

In conclusion, this work presented how efficiency was the segmentation and classification of agricultural scenes for navigation problem. As the data provided evince, this generated algorithm fulfils the expectations as far as segmenting is concerned, so that it sorts the appropriate classes. As a result, a modular strategy with ANN topology can be an option for the classification of segments with JSEG algorithm.

The tests performed with the orthogonal camera orientation to the crop rows showed that this is the best angle to extract individual objects of natural scenes. Through this angle was possible both to distinguish the objects captured by the images and get their characteristics for subsequent construction of the real object data model to be used in the virtual environment generation.

ACKNOWLEDGMENT

The authors would like to thanks to CAPES, CNPq and FINEP for funding this research project.

REFERENCES


