ELIMINATING CROP SHADOWS IN VIDEO SEQUENCES BY PROBABLE LEARNING PIXEL CLASSIFICATION

Tanghai Liu[1,*], Xiaoping Cheng[1]

[1] *Faculty of Computer and Information Science, Southwest University, Chongqing, China, 400715*

[*] *Corresponding author, Address: Faculty of Computer and Information Science, Southwest University, 1 Tiansheng Road, Beibei, Chongqing, 400715, P. R. China, Tel: +86-13647662870, Email: liutangh@yahoo.com.cn*

Abstract: Shadows have been one of the most serious problems for vegetation segmetation, especially under conditions of natural random airflow and human or vehicle disturbance. A video sequence processing method has developed in this paper to identify and eliminate crop shadows. The method comprises pixel models and algorithms explained in a probable learning framework. Expectation maximization (EM) for mixture models is established and an incremental EM method is proposed. This method performs a probable reasoning unsupervised classification of pixels for real-time implementation. The results show that the method is quite robust and can successfully remove shadows under natural lighting conditions.

Keywords: probable learning, shadows, vegetation segmentation, video processing

1.   INTRODUCTION

In earlier work of remote sensing technology as part of the agriculture related, it was shown that images acquired from satellites or aircrafts can provide effective help in agricultural related information such as woods, cropland, soil, and plant density.

Recently, precision agriculture (PA) is popular as a management strategy that uses information technologies to bring data from multiple sources to bear on decisions associated with agricultural production (NRC, 1997; Robert, 2002). In addition, PA needs low altitude images (Vioix et al., 2002), even images very close to examine crops (Shrestha et al., 2003), which have been acquired for analysis.

Advances in computer technology and digital video technology had increasingly opened up using video cameras in real-time field application for the economical acquisition of images. (Easton and Easton, 1996) developed a mechanical sensing system for counting young corn plants, which was mounted on a one-wheeled, human-powered cart. (Shrestha et al., 2003) developed a machine vision-based corn plant population sensing system to measure early growth stage corn population. Video was acquired from a vehicle-mounted digital video camera under different daylight conditions. (Steward et al., 1999) mounted a video camera at a height of 3.35 m on a custom-made camera boom and took video streams at a slow forward travel speed to keep the full horizontal resolution available in the video signal.

The next step after image acquisition was segmentation of vegetation from background. Different methods were available for separating vegetation from non-vegetation regions. (Pérez et al., 2000) used a normalized difference index (NDI) along with morphological operations for plant segmentation. (Tang et al., 2003) used Gabor wavelets and an artificial neural network for classification of broadleaf and grass weeds. (Steward et al., 2004) developed a method called reduced-dimension clustering (RDC) for vegetation segmentation. Besides, segmentation of monochrome field scene images were typically accomplished by thresholding the intensity histograms, which typically had bimodal distributions of pixel gray levels (Meyer et al., 1998; Andreasen et al., 1997).

Various vegetation segmentation researches, however, had little to do with shadows. Some researchers tried to avoid shadows by taking top down video (Steward et al., 2004; Steward et al., 1999; Shrestha et al., 2003; Vioix et al., 2002).

We learn that shadows are significant consideration in vegetation segmentation. Since sunshine cast is not always perpendicular, vegetation must have naturally shadows. Shadows pose serious difficulty in correct segmentation (see Figure 1). Eliminating shadows can improve qualities of segmentation. Also, with the help of shadows elimination approach, the field observation time can be prolonged. Although sunshine casts have usually angles, we need not await special sunshine.
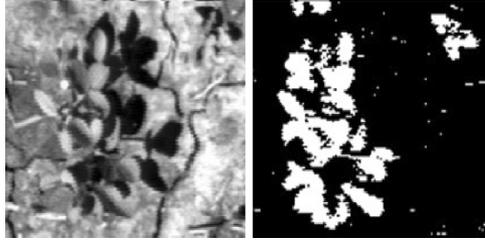
Figure 1. The left image has significant shadow components; obviously, the right image was segmented imprecisely because of existence of shadows. (Steward et al., 1999)

Furthermore, there are other natural conditions. For example, in the field, wind and breeze are common. At those cases when video acquisition processing needs human or vehicle participation, which should unavoidably disturb airflow, crop will swing. Then shadows are also not still.

Most researchers utilized video cameras for just acquiring still image frames, whose practices didn't exploit video camera capabilities fully. A video sequence provides various possibilities for improving field image processing. In fact, multiple temporal consecutive frames of images produced by video cameras can be used in removing noise, subtracting images from the background and so on. In this paper, we propose a video-processing-based approach to remove crop shadows. By using a probable reasoning method, our approach is quite robust and can successfully remove shadows under conditions of natural random airflow and human or vehicle disturbance.

"Background subtraction" is an old technique for finding moving objects in a video sequence. It succeeds not only in detecting moving object, but also their shadows. Each pixel must be classified before being used to update the background model. We show how this can be done properly, using a probabilistic classifier and a stable updating algorithm.

2. METHOD

2.1 Pixel model

Consider a single pixel and the distribution of its values over time. Some of the time it will be in its "normal" background state. Some of the time it may be in the shadow of swinging plants, and some of the time it may be part of a plant. Thus, we can think of the distribution of values $i_{x,y}$ of a pixel (x, y) as the weighted sum of three distributions $c_{x,y}$ (crop), $f_{x,y}$ (field), and $s_{x,y}$ (shadow):

$$i_{x,y} = w_{x,y} \cdot (c_{x,y}, f_{x,y}, s_{x,y})$$

The model for pixel (x, y) is parameterized by the parameters $\Theta = \{\omega_l, \mu_l, \Sigma_l : l \in \{c, f, s\}\}$ so that $w_{x,y} = (\omega_c, \omega_f, \omega_s)$, $f_{x,y} \sim N(\mu_f, \Sigma_f)$, and so on. For clarity, we omit the subscript x, y from the names of these parameters.

Let i be a pixel value. Let L be a random variable denoting the label of the pixel in this image. Our model defines the probability that $L = l$ and $I(x, y, t) = i$ to be:

$$P\big(L = l, I(x, y, t) = i \big| \Theta\big) = \omega_l \bullet (2\pi)^{-\frac{2}{d}} |\Sigma|^{-\frac{1}{2}} \exp\{-\tfrac{1}{2}(i - \mu_l)^T \sum\nolimits_l^{-1} (i - \mu_l)\}.$$

Where $I(x, y, t)$ is the instantaneous pixel value for the (x, y) pixel at time t.

Given these probabilities, we can classify the pixel value. Namely, we choose the class l with highest posterior probability $P(L = l \mid I(x, y, t))$.

2.2   Algorithms

2.2.1   EM for mixture models

Suppose we observe a sequence of pictures 1,…, T, and that $I(x, y, t)$ is the value of pixel (x, y) in the t-th image. We want to learn the parameters of the distribution $c_{x,y}$, $f_{x,y}$, and $s_{x,y}$, as well as the relative weights $w_{x,y}$. Then, we define the likelihood of a set of parameters Θ to be the probability of the data given Θ: $\prod_{t=1}^{T} P(L = l_t, I(x, y, t) \mid \Theta)$. We want to choose the parameters that maximize the likelihood. The sufficient statistics for this mixture estimation is N_l, M_l, and Z_l, where

N_l is the number of images for which $L_t = l$;

M_l is the sum of the input vectors, $\sum_{t=1,\dots,T, L_t = l} I(x, y, t)$; and

Z_l is given by $\frac{1}{T} \sum_{t=1,\dots,T, L_t = l} I(x, y, t) \cdot I(x, y, t)^T$, the sum of the outer products of the input vectors with themselves.

From these sufficient statistics, we can compute:

$$\omega_l = \frac{N_l}{\sum_l' N_l} \tag{2}$$

$$\mu_l = \frac{M_l}{N_l} \tag{3}$$

$$\Sigma_l = \frac{1}{N_l} Z_l - \mu_l^T \mu_l \qquad\qquad\qquad (4)$$

The likelihood with respect to the observable data is $L(\Theta) = \prod_{t=1}^{T} P(I(x,y,t) \mid \Theta)$. The EM algorithm explores a sequence of parameter setting, where each setting is found by using the previous one to classify the data. Formally, we compute the expected value of the sufficient statistics as follows:

$$E[N_l \mid \Theta^k] = \sum_{n=1}^{T} P(L_t = l \mid I(x,y,t), \Theta^k)$$

$$E[S_l \mid \Theta^k] = \sum_{n=1}^{T} P(L_t = l \mid I(x,y,t), \Theta^k) I(x,y,t)$$

$$E[Z_l \mid \Theta^k] = \sum_{n=1}^{T} P(L_t = l \mid I(x,y,t), \Theta^k) I(x,y,t) I(x,y,t)^T$$

Then defining Θ^{k+1} by using Equation 2-4 with the expected sufficient statistics.

2.2.2   Incremental EM

The standard EM procedure we just reviewed requires us to store the values of pixel (x, y) for all the images we have observed. This is clearly impractical for our application. We now describe an incremental variant of EM that does not require storing the data. This procedure was introduced by (Nowlan, 1991), and is best understood in terms of the results of (Neal and Hinton, 1993).

Neal and Hinton show that we can think of the EM process as continually adjusting the sufficient statistics. In this view, on each iteration when we process an instance, we remove its previous contribution to the sum and replace it with a new one. Thus, for example, when we update N_l, we remove $P(L_t = l \mid I(x,y,t), \Theta^{k'})$ and $P(L_t = l \mid I(x,y,t), \Theta^k)$, where $\Theta^{k'}$ are the parameter setting we used to compute the previous estimated statistics from $I(x,y,t)$, and Θ^k are the current parameter settings. Neal and Hinton show that after each instance is processed, the likelihood of the data increases. Whenever we observe a new instance, we add its contribution to the sufficient statistics. Thus, in the long run, this process converges to a local maximum with high probability.

The resulting procedure for each pixel (x, y) has the following structure:
Initialize parameters Θ.

$t \leftarrow 0$

for $l \in \{c, f, s\}$

$N_l \leftarrow k\omega_l$

$S_l \leftarrow k\omega_l \cdot \mu_l$

$Z_l \leftarrow k\omega_l \Sigma_l + \mu_l \cdot \mu_l^T$

Loop the following

$t \leftarrow t + 1$

for $l \in \{c, f, s\}$

$\quad N_l \leftarrow N_l + P(L_t = l \mid I(x, y, t), \Theta)$

$\quad S_l \leftarrow S_l + P(L_t = l \mid I(x, y, t), \Theta) I(x, y, t)$

$\quad Z_l \leftarrow Z_l + P(L_t = l \mid I(x, y, t), \Theta) I(x, y, t) I(x, y, t)^T$

Compute Θ from $\{N_l, S_l, Z_l\}$.

The initialization step of the procedure sets the statistics to be the expected statistics for the initial choice of Θ. Then, in each iteration we add the expected statistics for the new instance to the accumulated statistics.

3.   EXPERIMENTS AND CONCLUSIONS

For testing the performance of the method for eliminating shadows, we take several continuous video streams of maize crops. Figure 2 shows some typical frames.

In order to distinguish shadows from vegetation, pixel models and incremental EM are determined. The related results are shown in Figure 3.

It can be seen from following examples; our real-time approach has quite good shadows elimination capability, which paves the way for further processing in various field applications. Also, our method has consistently high accuracy in different video streams, which means robustness desired in natural condition.

However, our approach cannot deal with situation that shadows, which vegetation cast, cover vegetation. Our algorithm could be further integrated to eliminate shadows for robustness enhancement and possible accuracy improvement. We are quite sure our shadows identification method is a promising pre-processing technique for various agricultural applications.

Figure 2. Frames from video sequences in which crop cast shadows

Figure 3. Shadows casting on soil are eliminated via our approach

ACKNOWLEGEMENTS

REFERENCES

Andreasen, C., M. Rudemo, and S. Sevestre, Assessment of weed density at an early stage by use of image processing, *Weed Research* 37, 1997: 5–18.

Easton, D. T., and Easton, D. J, Device to measure and provide data for plant population and spacing variability, U. S. Patent No. 5, 568, 405, 1996.

Jean-Baptiste Vioix, Jean-Paul Douzals, Frédéric Truchetet, Louis Assémat, and Jean-Philippe Guillemin, Spatial and spectral methods for weed detection and Localization, *EURASIP Journal on Applied Signal Processing* 2002, 7: 697–685.

Meyer, G. E., T. Mehta, M. F. Kocher, D. A. Mortensen, and A. Samal, Textural imaging and discriminant analysis for distinguishing weeds for spot spraying, *Trans. ASAE* 1998, 41 (4): 1189–1197.

National Research Council, *Geospatial and Information Technologies in Crop Management*, Washington, D. C. National Academy Press, 1997.

Neal, R. M., and G. E. Hinton, A new view of the EM algorithm that justifies incremental and other variants, Unpublished manuscript, 1993.

Nowlan, S. J., *Soft Competitive Adaptation: Neural Network Learning Algorithms based on Fitting Statistical Mixtures*, PhD thesis, School of Computer Science, Carnegie Mellon University, 1991.

Pérez, A. J., F. López, J. V. Benlloch, and S. Christensen, Colour and shape analysis techniques for weed detection in cereal fields, *Computers and Electronics in Agriculture* 2000, 25 (3): 197–212.

Robert, P. C, Precision agriculture: a challenge for crop nutrition management, *Plant and Soil* 2002, 247 (1): 143–149.

Shrestha, D. S., and B. L. Strward, Automatic corn plant population measurement using machine vision, *Transactions of the ASAE* 2003, 46 (2): 559–565.

Steward, B. L., and L. F. Tian, Machine-vision weed density estimation for real-time outdoor lighting conditions, *Transactions of the ASAE*, 1999, 42 (6): 1897–1909.

Steward, B. L., L. F. Tian, D. Nettleton, and L. Tang, Reducee-dimension clustering for vegetation segmentation, *Transactions of the ASAE*, 2004, 47 (6): 609–616.

Tang, L., L. Tian, and B. L. Steward, Classification of broadleaf and grass weeds using gabor wavelets and an artificial neural network, *Transactions of the ASEA*, 2003, 46 (4): 1247–1254.