# Shape-based Segmentation of Tomatoes for Agriculture Monitoring

Ujjwal Verma[1,2], Florence Rossant[1], Isabelle Bloch[2], Julien Orensanz[3] and Denis Boisgontier[3]

[1]*ISEP, Paris, France*
[2]*Institut Mines-Telecom, Télécom ParisTech, CNRS LTCI, Paris, France*
[3]*Cap2020, Gironville sur Essonne, France*
{*ujjwal.verma, isabelle.bloch*}*@telecom-paristech.fr, florence.rossant@isep.fr,* {*j.orensanz, d.boisgontier*}*@cap2020.fr*

Abstract:     In this paper, we present a segmentation procedure based on a parametric active contour with shape constraint, in order to follow the growth of the tomatoes from the images acquired in the field. This is a challenging task because of the poor contrast in the images and the occlusions by the vegetation. In our sequential approach, considering one image per day, we assume that a segmentation of the tomatoes is available for the image acquired the previous day. An initial curve for the active contour model is computed by combining gradient information and region information. Then, an active contour with shape constraint is applied to provide an elliptic approximation of the tomato boundary. We performed a quantitative evaluation of our approach by comparing the results with the manual segmentation. Given the varying degree of occlusion in the images, the image data set was divided into three categories, based on the occlusion degree of the tomato in the processed image. For the cases with low occlusion, good results were obtained, with an average relative distance between the manual segmentation and the automatic segmentation of 2.73% (expressed as percentage of the size of tomato). For the images with significant amount of occlusion, a good segmentation was obtained on 44% of the images, where the average error was less than 10%.

## 1  INTRODUCTION

Optimal harvesting date and predicted yield are valuable information when farming open field tomatoes, making harvest planning and work at the processing plant much easier. Monitoring tomatoes during their early stages of growth is also interesting to assess plant stress or abnormal development. Satellite data and crop growth modeling are generally used for estimating the yield of a large region (Prasad et al., 2006),(Mkhabela et al., 2011). However, satellite data are affected by adverse climatic conditions (clouds, etc.) resulting in inaccurate predictions(Mkhabela et al., 2011). Crop growth modeling, which integrates information regarding the cultivated plant, soil and weather conditions, considers the ideal case with no infected plant. Recent studies have concentrated on combining these two approaches(Zhao and Pei, 2013). Nevertheless these methods depend on the quality of the different parameters involved (vegetation indices, soil and weather information) and they are not accurate enough to detect abnormal development.

In this work, we present a different approach where we intend to monitor the growth of tomatoes and measure their size in an open field. For this purpose, two cameras are installed in the field and two images are captured at regular intervals. In order to avoid a complete 3D reconstruction, we assume that a tomato can be approximated by a sphere in the 3D space, which projects into an ellipse in the image plane. Hence, the first part of our system aims at detecting and segmenting the tomatoes in both images, using elliptic approximations. Then, the second part aims at estimating the sphere radius, using the camera parameters. An estimate of the yield is obtained from this information. In this paper, we focus on the segmentation procedure only.

Computer vision algorithms have been applied in the agricultural domain in order to replace human operators with an automated system. They have been used to grade and sort agricultural products (Jayas et al., 2000; Du and Sun, 2006; Narendra and Hareesh, 2010), to detect weeds in a field (Aitkenhead et al., 2003; Yang et al., 2000; Lee et al., 1999), and to model the growth of fruits and then predict the

yield (Aggelopoulou et al., 2011; Stajnko and Cmelik, 2005). In (Aggelopoulou et al., 2011), the yield of an apple orchard is estimated using only the density of flowers. In (Stajnko and Cmelik, 2005), only 5 images captured at different stages of the apple maturation are studied in order to predict the yield. However, to the best of our knowledge, there has not been any related work where the growth of a fruit or vegetable is studied based on the images captured during the entire agricultural season.

There is little growth of a tomato during a given day. So, only one image per day is analyzed in this work, thus creating a series of images (approximately 20-30 images). One of the difficulties of the segmentation part is occlusion: most of the tomatoes are partially hidden by other tomatoes and/or leaves (Figure 1). Moreover, color information is not of much use as tomatoes are red only at the end of the ripening. Also, another difficulty is a very low contrast in some cases due to shadows.
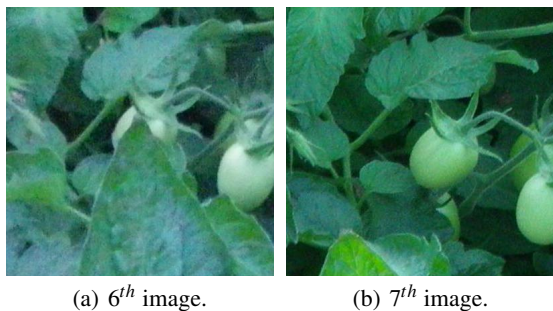


(a) $6^{th}$ image.  (b) $7^{th}$ image.

Figure 1: Two successive images of the tomato $S = 7$.

In this work, the segmentation should be as automatic as possible. However, we assume that an operator validates each obtained segmentation. If the result is poor, the operator rejects it. Indeed, given the difficulties, the segmentation is a very challenging task, and a manual validation is preferable. This approach enables us to use the segmentation done in the $i^{th}$ image (if validated) as a reference for the segmentation of the same tomato in the $(i+1)^{th}$ image.

In order to segment the tomatoes, we use a parametric active contour model, which allows us to introduce *a priori* knowledge on the shape of the object to be segmented, thus making the segmentation more robust to noise and occlusion. Using an elliptic shape constraint is consistent with our prior assumption.

The main steps of the segmentation algorithm are the following: first, gradient information is used in order to find the candidate contour points and propose several elliptic approximations using the RANSAC algorithm. Secondly, region information is added, enabling us to select the best ellipse for the initialization of the active contour and finding the regions of potential occlusions. Third, the active contour with elliptic constraint is applied. Finally, four ellipse estimates are computed. The operator has only to select the best one as the final segmentation.

The original features of the proposed algorithm include the approximation of the tomatoes as ellipses and the conditioning of the computation of the image energy by the non-occluded regions. These features allow coping with occlusions and local loss of contour and edges.

We present the active contour model with shape constraint in Section 2 and the different steps of the segmentation algorithm in Section 3. Section 4 discusses the experimental results.

## 2 ACTIVE CONTOUR WITH AN ELLIPTICAL SHAPE PRIOR

Parametric active contour model or snake was originally introduced by (Kass et al., 1988) in order to detect a boundary of an object in an image. This algorithm deforms the contour iteratively from its initial position towards the edges of an object by minimizing an energy functional. The energy functional associated with the contour $v$ is usually composed of three terms:

$$\mathbf{E}_T(v) = \mathbf{E}_I(v) + \mathbf{E}_{Im}(v) + \mathbf{E}_{Ext}(v) \quad (1)$$

where $\mathbf{E}_I(v)$ is the internal energy controlling the smoothness of the curve and $\mathbf{E}_{Im}(v)$ is the energy derived from image data. The external energy $\mathbf{E}_{Ext}(v)$ can express contextual information, such as shape information. The authors in (Foulonneau et al., 2006) used Legendre moments to define an affine invariant shape prior in a region based active contour. In our case, the region information is not significant due to the presence of leaves (occlusions) and other tomatoes of similar intensity profile. In (Charmi et al., 2009), Fourier descriptors are used in order to align the active contour with the reference curve of suitable shape and orientation. In our work, the tomato in each image is assumed to have an ellipse shape. Since ellipses are easily represented in a parametric form from a few parameters, it is natural to propose a parametric active contour model.

Let us define the reference ellipse as $z_e$. This ellipse is estimated from the evolving contour $z$. Both curves are expressed in polar coordinates with the origin at the center of $z_e$:

$$z_e(\theta) = r_e(\theta)e^{j\theta}, z(\theta) = r(\theta)e^{j\theta}, \theta \in [0, 2\pi] \quad (2)$$

Our energy functional with an elliptic shape regularization is defined as:

$$\mathbf{E}_T(r, r_e) = \int_0^{2\pi} \frac{\alpha}{2}|r'(\theta)|^2 d\theta + \int_0^{2\pi} E_{Im}(r(\theta)e^{j\theta})d\theta$$
$$+ \frac{\psi}{2}\int_0^{2\pi}|r(\theta) - r_e(\theta)|^2 d\theta \qquad (3)$$

In the above equation, the first term represents the internal energy which controls the variations of $r$ and makes it regular. The second term is a classical image energy calculated from the gradient vector flow (Xu and Prince, 1998). The last term restricts the evolving contour to be close to the reference ellipse. The parameter $\alpha$ controls the smoothness of the curve, and $\psi$ controls the influence of the shape prior on the total energy. Note that instead of modifying a 2-D vector $v(s) = (x(s), y(s))$ as in the classical active contour model, only a 1-D vector $r(\theta)$ is modified for each value of the parameter $\theta$. Moreover, the shape constraint makes the usual second derivative term in the internal energy useless, and is therefore not included in the proposed energy functional.

The minimum of $\mathbf{E}_T$ is obtained in two steps: First, a least square estimate of the ellipse $z_e$ is computed from the initial contour $z_0$. Then, the evolving contour $z$ is computed by minimizing $\mathbf{E}_T$ while assuming $z_e$ fixed. From the evolving contour $z$ so obtained, the parameters of the least square estimate of the ellipse $z_e$ are regularly updated. This two-step iterative process is repeated, in order to obtain the minimum of $\mathbf{E}_T$.

The minimization of $\mathbf{E}_T$ with respect to $r$ is equivalent to solving the following Euler equation:

$$-\alpha r''(\theta) + \nabla E_{Im}(\theta) \cdot n(\theta) + \psi(r(\theta) - r_e(\theta)) = 0 \quad (4)$$

where $n(\theta) = [\cos\theta, \sin\theta]^T$.

To find iteratively a solution of this equation, we introduce a time variable, and the resulting equation is discretized using finite differences, as in the case of the classical active contours.

# 3  DETAILED ALGORITHM

In this section, we present an algorithm which allows us to follow the growth of a tomato, which has been manually segmented in the first image ($i = 1$).

Let us denote by $im^{i+1}$ the $(i+1)^{th}$ image of the tomato $S$. In the rest of this paper, an ellipse centered at $[xc, yc]$, whose semi major and minor axes lengths are $a$ and $b$, respectively, and which has a rotation angle of $\varphi$, is represented as $Ell = [xc, yc, a, b, \varphi]$. The tomato approximated by an ellipse in $im^i$ is represented as $Ell^i = [xc^i, yc^i, a^i, b^i, \varphi^i]$. In our sequential approach, the computation of the contour in the

$(i+1)^{th}$ image is based on both the information in $im^{i+1}$ and the contour of the tomato in the $i^{th}$ image. In the following sections, it is assumed that there is little movement and little growth of tomatoes between two successive images (Figure 2).

## 3.1  Pre-processing

As mentioned above, the color information is not of much use. However, the edges of tomatoes are more prominent in the red component of the image, and hence only this component is considered. The original image is cropped around the position $(xc^i, yc^i)$, resulting in a smaller image ($imS_c^{i+1}$). The contrast is enhanced by a contrast stretching transformation.

## 3.2  Updating the Tomato Position

Due to its increasing weight, the tomato tends to fall towards the ground (Figure 2) and its position in $imS_c^{i+1}$ is calculated, using pattern matching. The bright areas, that may correspond to the tomato, are extracted by convolving the cropped image with a binary mask representing a white disk of radius $\chi r^i$ where

$$r^i = \frac{a^i + b^i}{2} \qquad (5)$$

and $\chi$ is a constant determined experimentally. The local maxima $C_c^{i+1} = \{(x_k, y_k), k = 1, ..., k_n\}$ are then extracted. From these $k_n$ points, the one $C_m = (x_m, y_m)$, which is the closest to $(xc^i, yc^i)$, is selected as the new location of the tomato center (Figure 2(b)).

A new cropped image $imS^{i+1}$ is then extracted from $im^{i+1}$, centered at $C_m = (x_m, y_m)$. The size of this new image is adapted to the size of the tomato (deduced from $a^i$ and $b^i$) so that we restrict the region to be analyzed as much as possible, reducing thus the processing cost of the next steps. The contrast stretching transformation is applied to $imS^{i+1}$.



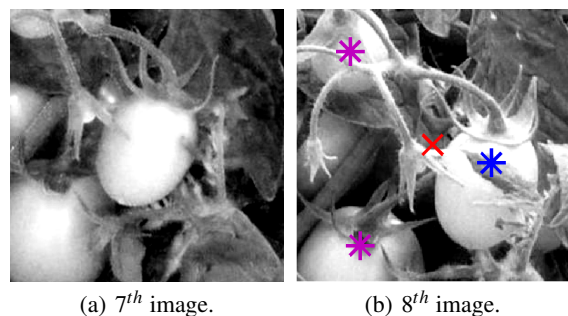(a) $7^{th}$ image.　　(b) $8^{th}$ image.

Figure 2: Updating the position of the tomato: previous position $(xc^i, yc^i)$ in red, candidate positions $C_c^{i+1}$ in magenta and blue, and new position $C_m$ in blue.

## 3.3 Elliptic Approximations

In order to obtain an initial contour for the active contour model, we first compute $l_n$ points which may lie on the boundary of the tomato. From these $l_n$ points, a RANSAC estimate is used to obtain several candidate ellipses. Finally, one of these ellipses is selected as the initial contour based on additional region information and size regularization (Section 3.4.2).

### 3.3.1 Detection of Tomato Contour Points

Let us take $C_m$ as the origin of the polar coordinate system. Then we select $l_n$ points $P_l = p_l e^{j\theta_l}$, where $l = 1,...,l_n$, $0 < \theta_l < 2\pi$, that satisfy the following three conditions:

$$0.5r^i < p_l < 1.5r^i \qquad (6)$$

$$|\arg(\nabla imS^{i+1}(P_l)) - \theta_l| \leq \frac{\pi}{8} \qquad (7)$$

$$|\nabla imS^{i+1}(P_l)| > \eta \qquad (8)$$

where $\nabla imS^{i+1}(P_l)$ is the gradient at $P_l$ in $imS^{i+1}$. The above conditions select the points of strong gradient whose direction is within an acceptable limit with respect to the vector normal to the circle with radius $r^i$. The threshold values have been set experimentally. As shown in Figure 5(a), most points lying on the boundary of the tomato have been correctly detected along with some additional points lying on the leaves.

### 3.3.2 Calculation of Several Ellipses from the Points $P_l$ using RANSAC Estimation

A least square estimate of an ellipse calculated from *all* $l_n$ points might result in a contour far away from the actual boundary because of the detection of irrelevant points. Therefore, we use a RANSAC (Fischler and Bolles, 1981) estimate based on an elliptic model in order to compute several candidate ellipses.

Under normal circumstances, the size and the orientation of the tomato in $imS^{i+1}$ are supposed to be close to the ones in $imS^i$. This information is incorporated in the RANSAC estimation and only the ellipses whose parameters satisfy the following conditions are considered:

$$-0.1 < \frac{a^{i+1} - a^i}{a^i} < 0.2, -0.1 < \frac{b^{i+1} - b^i}{b^i} < 0.2 \quad (9)$$

$$-0.1 < \frac{SA^{i+1} - SA^i}{SA^i} < 0.25 \quad (10)$$

$$|\frac{Ecc^{i+1} - Ecc^i}{Ecc^i}| < 0.1 \quad (11)$$

$$|\frac{\varphi^{i+1} - \varphi^i}{\varphi^i}| < 0.2 \quad (12)$$

where $SA^{i+1}$ and $SA^i$ represent the surface of the ellipses in $imS^{i+1}$ and $imS^i$ respectively. The eccentricity ($Ecc = \frac{a}{b}$) for the two ellipses is denoted by $Ecc^{i+1}$ and $Ecc^i$ respectively.

Negative variations for $a$ and $b$ (Equation 9) are possible because of the movement of the tomato with respect to the camera or because of the variation in the orientation, as tomatoes are actually not perfect spherical objects. Equation 10 restricts the apparent size of the tomato while Equation 11 restricts the admissible values for eccentricity, thus controlling the apparent shape of the tomato.

The threshold values in Equations 9-12 have been determined after studying the parameters of the ellipses obtained from the manual segmentation of five tomatoes. For example, Figures 3 and 4 shows the relative evolution of the surface $SA$ and length of semi-major axis $a$ of the ellipses. Most of the measurements are situated within the limits defined above. Note that the dissymmetry in the lower and upper bounds in Equations 9-10 is due to the fact that tomatoes are supposed to grow during the agricultural season.
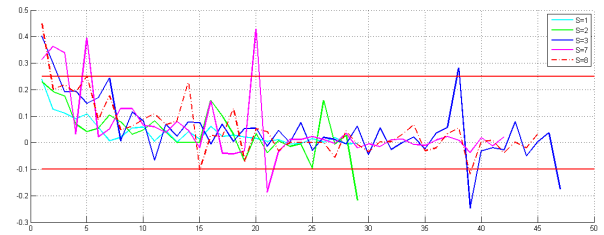


Figure 3: Evolution of $SA$. The abscissa represents the image number ($i$), and the ordinate represents $\frac{SA^{i+1} - SA^i}{SA^i}$. The solid horizontal red lines show the selected threshold values.
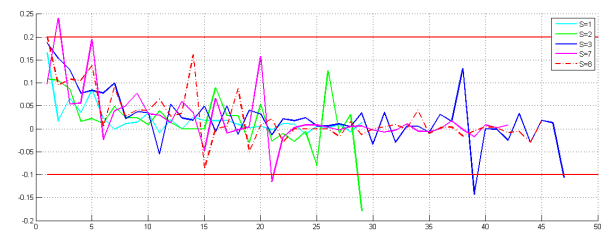


Figure 4: Evolution of $a$. The abscissa represents the image number ($i$), and the ordinate represents $\frac{a^{i+1} - a^i}{a^i}$. The solid horizontal red lines show the selected threshold values.

From the $N$ ellipses computed using the RANSAC algorithm, a total of $N_a$ ellipses, with $N_a < N$, are retained, corresponding to the $N_a$ ellipses with the largest number of inliers (Figure 5(b)).

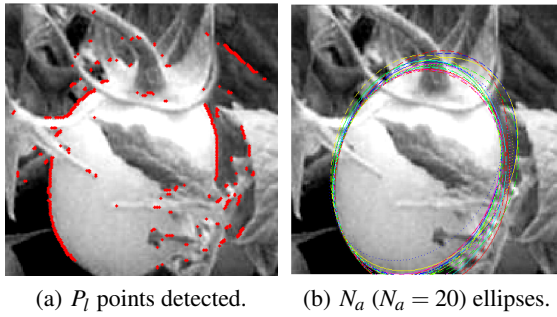(a) $P_l$ points detected.     (b) $N_a$ ($N_a = 20$) ellipses.

Figure 5: Points of strong gradient and ellipses detected using the RANSAC estimate.

## 3.4 Adding Region Information

A region growing algorithm is applied in order to add region information and determine the best initialization for the active contour among the $N_a$ ellipses $Ell_u^{i+1}$, where $u = 1, ..., N_a$. Moreover, potential occlusions are also derived from this information.

### 3.4.1 Using Region Growing to Calculate the Region Representing a Tomato

Let us denote by $\omega_u$ the binary image representing the region inside the ellipse $Ell_u^{i+1}$. We apply a classical region growing algorithm starting from $\omega_{seed}$ and limiting the growing to $\omega_{limit}$, where:

$$\omega_{seed} = \bigcap_{u=1}^{N_a} \omega_u, \quad \omega_{limit} = \bigcup_{u=1}^{N_a} \omega_u \qquad (13)$$

The final region is denoted by $\omega_t$ (Figure 6(a)).

### 3.4.2 Selecting the Initial Contour

We define $\tau_m$ as:

$$\tau_m = \min_{u=1,2,....N_a} \tau(u) \qquad (14)$$

with

$$\tau(u) = \frac{|\omega_u \cap (1-\omega_t)| + |(1-\omega_u) \cap \omega_t|}{|\omega_u \cap \omega_t|} \qquad (15)$$

where $|A|$ represents the cardinality of a set $A$. $\tau(u)$ measures the consistency between the segmentation obtained through the contour analysis $\omega_u$ and the region analysis $\omega_t$. It reaches a minimum (zero) when $\omega_u$ and $\omega_t$ match perfectly.

Let us denote by $a_u^{i+1}$ and $b_u^{i+1}$ the axes lengths of the candidate ellipse $Ell_u^{i+1}, u \in [1, N_a]$. We select the ellipse $v$ (Figure 6(b)) that minimizes $\left[(a_u^{i+1} - a^i)^2 + (b_u^{i+1} - b^i)^2\right]$ under the condition

$$\tau(v) \leq 1.1 \, \tau_m \qquad (16)$$

Thus, we have obtained the initial contour by combining the results obtained using two different segmentation methods, one based on boundary information and the other based on region information. The selected ellipse $Ell_v^{i+1}$ is chosen among the ones for which both results are consistent, allowing therefore a better robustness with respect to occlusions. Moreover, another regularization condition is added, which imposes that the size and shape of the ellipse in $imS^{i+1}$ are close to the ones in $imS^i$.
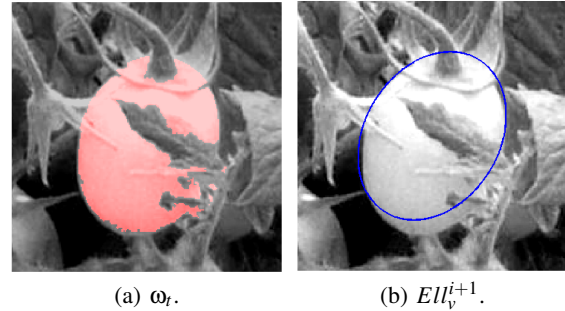


(a) $\omega_t$.     (b) $Ell_v^{i+1}$.

Figure 6: Result of the region growing ($\omega_t$) and selection of the initial ellipse $Ell_v^{i+1}$.

### 3.4.3 Finding Potential Occlusion

This part aims at finding the regions where occlusions could disturb the behavior of the active contour. For example, the region in which the tomato is attached to the plant has a different intensity from the one of the tomato.

Let $Ell_{te}$ denote the ellipse which covers the convex hull of $\omega_t$ and which minimizes the number of pixels inside the ellipse $Ell_{te}$ and not belonging to the region $\omega_t$ (Figure 7(a)). Let $\omega_{te}$ be the region inside $Ell_{te}$. Then, the region of occlusion $\omega_{oc}$ can be computed as:

$$\omega_{oc} = \omega_{te} \cap \omega_t^c \qquad (17)$$

Using morphological operations (erosion followed by reconstruction by dilation), small regions are removed from $\omega_{oc}$, so that the resulting $\omega_{oc}$ corresponds to actual leaves causing the occlusions (Figure 7(b)). Apart from detecting the "head" of the tomato, any other additional occlusion (mostly due to leaves) can also be detected using this approach (Figure 7(b)).

## 3.5 Applying Active Contours

The active contour (Section 2) is applied with the following initialization $Ell_{vc}^{i+1} = [xc_v^{i+1}, yc_v^{i+1}, 0.95a_v^{i+1}, 0.95b_v^{i+1}, \varphi_v^{i+1}]$. Indeed, the movement of the curve $z$ is smoother and faster if initialized inside the tomato. For the first $n_{start}$
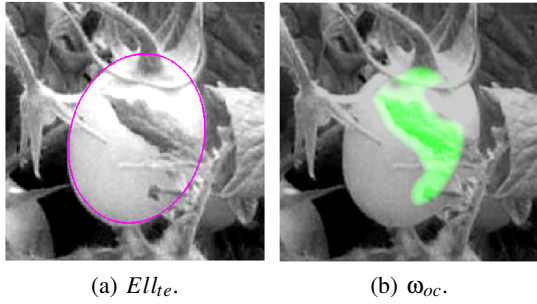
(a) $Ell_{te}$.      (b) $\omega_{oc}$.

Figure 7: Detecting the regions of potential occlusion.

iterations, the parameter $\psi$ is set to zero, so that $z$ moves towards the most prominent contours. Then the shape constraint is introduced for $n_{ellipse}$ iterations ($\psi \neq 0$) in order to guarantee robustness with respect to occlusion. Finally, the shape constraint is relaxed ($\psi = 0$) for a few $n_{end}$ iterations, which guarantees reaching the boundary more accurately, as a tomato is not a perfect ellipse.

Note that the image forces are not considered in the region of occlusion $\omega_{oc}$, in every step of this process.

**Updating the Reference Ellipse:** As explained in Section 2, the reference ellipse $z_e$ is updated every $n_{shape}$ iterations. A least square estimate calculated from all the points of the curve $z$ is not relevant, because some of them may lie on false contours (e.g. leaves). So, the following algorithm aims at selecting a subset of points that actually lie on the boundary of the tomato.

We use a polar coordinate system with the origin at the center of the current reference ellipse $z_e$. As in Section 2, let us denote by $z(\theta) = r(\theta)e^{j\theta}$ a point of the evolving curve, $z_e(\theta) = r_e(\theta)e^{j\theta}$ the corresponding point on the reference ellipse, $n_e(\theta)$ the vector normal to the ellipse $z_e$, and $z_q(\theta) = r_q(\theta)e^{j\theta}$ the point that maximizes the gradient module for $0 < r_q(\theta) < 1.1 r_e(\theta)$. The point $z(\theta)$ is selected as a point lying on the boundary of the tomato if it satisfies the following conditions:

$$|\nabla imS^{i+1}(z(\theta)) \cdot n_e(\theta)| > \Gamma \qquad (18)$$

$$\frac{|\nabla imS^{i+1}(z(\theta)) \cdot n_e(\theta)|}{|\nabla imS^{i+1}(z(\theta))|} > 0.75 \qquad (19)$$

$$d(z_q(\theta), z(\theta)) < d_{max} \qquad (20)$$

where $\cdot$ represents the vector dot product.

The first condition ensures that the magnitude of the gradient vector projected onto the normal of the ellipse is strong. The threshold $\Gamma$ is determined automatically (Otsu, 1975). The second condition ensures that the direction of the gradient is close to the vector

normal to the ellipse. The last condition ($d_{max} = 2$ in our experiments) imposes that the point considered is a meaningful local maximum of the gradient. Finally, the parameters of the reference ellipse are updated by calculating a least square approximation from the subset of points lying on the evolving contour $z$ selected using the above conditions (Figure 8(a)).
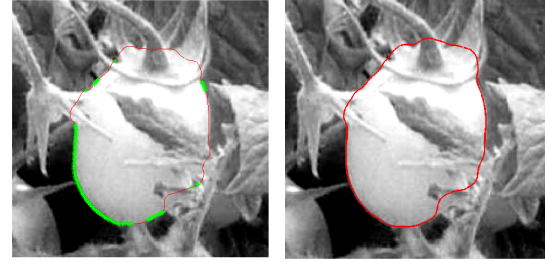


(a) Updating $z_e$ from selected points (green).    (b) Final contour $z$.

Figure 8: Active contour with shape constraint.

## 3.6 Refining the Results

A least square estimate of an ellipse from $z$ (Figure 8(b)) is generally not relevant as outliers may be present due to occlusion. So, again, a selection procedure is applied. A first subset of points $\mathbf{P}_h$ is obtained by using criteria similar to the ones described in Section 3.5 (Equations 18-20) . Then, another subset $\mathbf{P}'_h$ is computed by relaxing the condition related to the gradient direction (Figure 9).
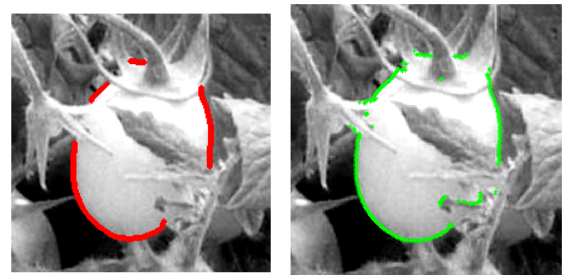


(a) set of points $\mathbf{P}_h$.    (b) Points $\mathbf{P}'_h$.

Figure 9: Two different set of points $\mathbf{P}_h$ and $\mathbf{P}'_h$.

**Final Contour:** Then four ellipses are computed as follows:

1. A least square approximation $Ell_{f1}^{i+1} = [xc_{f1}^{i+1}, yc_{f1}^{i+1}, a_{f1}^{i+1}, b_{f1}^{i+1}, \varphi_{f1}^{i+1}]$ is computed from all the points of $z$.

2. Another estimate $Ell_{f2}^{i+1} = [xc_{f2}^{i+1}, yc_{f2}^{i+1}, a_{f2}^{i+1}, b_{f2}^{i+1}, \varphi_{f2}^{i+1}]$ is obtained from $\mathbf{P}'_h$

using the RANSAC algorithm with the following conditions:

$$0.9a_{f1}^{i+1} < a_{f2}^{i+1} < 1.1a_{f1}^{i+1} \qquad (21)$$

$$0.9b_{f1}^{i+1} < b_{f2}^{i+1} < 1.1b_{f1}^{i+1} \qquad (22)$$

3. A least square approximation $Ell_{f3}^{i+1}$ is obtained from the subset $\mathbf{P}_h$.

4. A weighted least square estimate $Ell_{f4}^{i+1}$ is obtained where the points of $\mathbf{P}_h$ are assigned a higher weight (0.75) and the other points of $z$ a lower weight (0.25). This is done in order to give importance to the points that are surely on the boundary of the tomato.

If the images have a good contrast, and little or no occlusion, all the four ellipses will be almost identical (Figure 10(a)). However, in case of occlusions and poor contrast, the four ellipses may be different (Figure 10(b)).



(a) $8^{th}$ image.   (b) $17^{th}$ image: $\mathbf{P}_h$ (yellow) and $\mathbf{P}_h'$ (green).
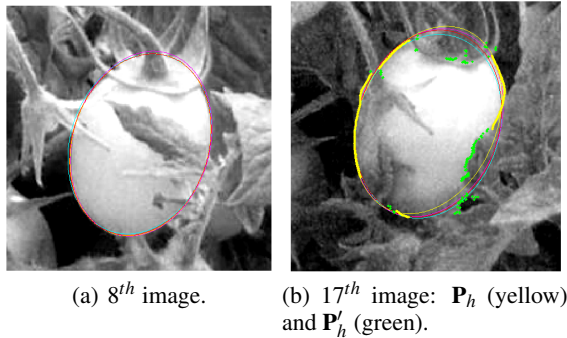
Figure 10: Final ellipse estimates for two different images.

## 4 RESULTS

Two cameras (Pentax Optio W80) were installed in an open field of tomatoes. The same setup was used for two agricultural seasons (April-August, 2011 and 2012), capturing one image per day. We have identified 10 tomatoes, covering different sites and different seasons, thus ensuring variability (302 images in total). The tomatoes were identified manually by observing the images of the entire agricultural season. Due to the severe occlusions, only a limited number of tomatoes were visible in most of the images of a given season. Therefore, only the tomatoes which were visible in more than 10 consecutive images were studied.

The obtained segmentations $A$ were compared with the manual segmentations $M$ (approximated by ellipses) by computing the average $D_{mean}^i$ and maximal $D_{max}^i$ distance between $A$ and $M$ for the $i^{th}$ image (expressed in pixels). In order to better interpret

the results, the maximum and mean distances between two contours are normalized by the size of the tomato as:

$$D_{meanR}^i = \frac{D_{mean}^i}{r^i}100 \qquad (23)$$

$$D_{maxR}^i = \frac{D_{max}^i}{r^i}100 \qquad (24)$$

where $r^i$ is defined in Equation 5.

The data set contains images of varying contrast and degree of occlusions. Obviously, it is impossible to obtain a reliable segmentation in case of severe occlusion, even manually. Consequently we studied experimentally the effect of the percentage of occlusion on the final estimation of the radius of a spherical object (considering the complete system, segmentation and partial 3D reconstruction). In our experiments, the percentage $P$ of occlusion corresponds to the occlusion of an arc with subtended angle equal to $\frac{2\pi P}{100}$. For less than 30% occlusion, the variation in the estimated radius was very small, and for more than 30% occlusion, significant change in the values of the estimated radius was observed. Thus, we identified three different categories:

- Category 1, containing images with an amount of occlusion $P$ less than 30% for which the estimation is very robust with respect to segmentation imprecision (Figure 11(a)),

- Category 2, with $30\% < P < 50\%$ which is more prone to segmentation error (Figure 11(b)),

- Category 3, with $P > 50\%$ for which it is impossible to perform a reliable segmentation (Figure 11(c)).

The percentage of occlusion was determined manually by selecting the end points of the occluded elliptic arc. Note that the percentage of occlusion was computed only to evaluate the segmentation procedure, and this is not a part of the algorithm.

Table 1 shows the mean ($\mu$) and the standard deviation ($\sigma$) for $D_{meanR}$ and $D_{maxR}$, for the images belonging to category 1. Good results were obtained even in the presence of occlusion by nearby leaves/branches and tomatoes (Figures 12(a) and 12(b)). Also, a low $\mu_{D_{meanR}}$ along with lower $\sigma_{D_{meanR}}$ demonstrates the robustness of the proposed method. However, in some images captured at the beginning of the season, when the size is very small, the occlusion due to leaves present on the "head" of the tomato results in an ambiguity on the position of the actual contour (Figure 12(c)). Also, in some images (Figure 12(d)), due to a shadow effect on a portion of the contour, the intensity profile of the tomato and the adjacent leaves are nearly identical, resulting in a very low contrast. Such cases may result in
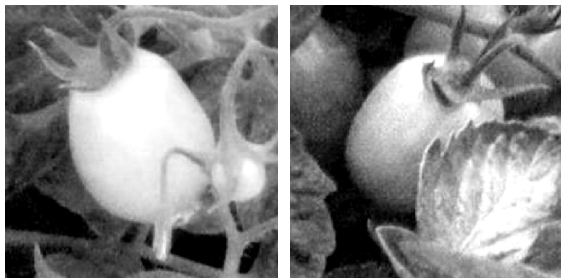
(a) Category 1: Tomato $S = 1$, Image 8.
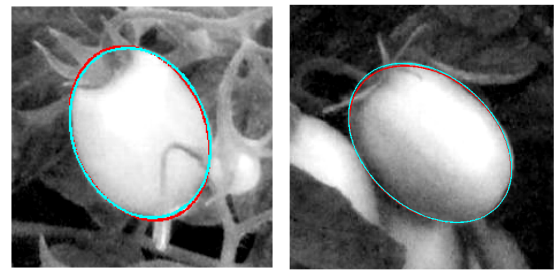(b) Category 2: Tomato $S = 7$, Image 17.



(c) Category 3: Tomato $S = 2$, Image 14.

Figure 11: Three different categories of image based on the occlusion.



(a) Tomato $S = 1$, Image 8 (1.89%, 4.70%).
(b) Tomato $S = 4$, Image 24 (1.29%, 4.39%).

(c) Tomato $S = 5$, Image 4 (4.71%, 11.84%).
(d) Tomato $S = 8$, Image 5 (4.77%, 16.04%).

Figure 12: Final segmentation $Ell_{f4}$ (red) obtained on images of category 1. The contour in cyan represents the manual segmentation. Also shown are the distance measures. The values in brackets are ($D_{meanR}$, $D_{maxR}$).



(a) Tomato $S = 7$, Image 17 (1.94%, 5.45%).
(b) Tomato $S = 4$, Image 18 (2.22%, 6.16%).

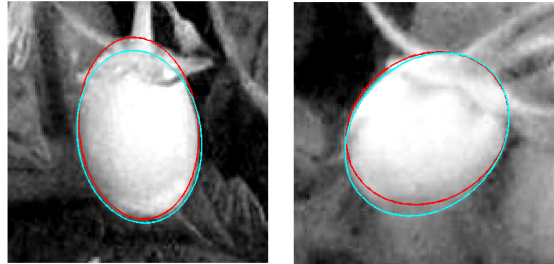Figure 13: Final segmentation $Ell_{f4}$ (red) obtained on images of category 2.

comparatively high distance measures even in the absence of any occlusion.

In the images of category 2, containing a significant amount of occlusion, good results were obtained on 44 % of the images (Figure 13), where $D_{meanR} < 10\%$. For this category $\mu_{D_{meanR}}$ lies in the interval $[2,30]\%$ for ellipse $Ell_{f4}$. The significantly higher distance measure for some images of this category is caused by the false detection of the position of the tomato due to the occlusion. The position of the tomatoes was not correctly detected in approximately 6% of the images belonging to category 2. Note that $Ell_{f4}$ is not necessarily the best ellipse, and is selected for illustrative purpose only.

Due to the variation in the contrast and occlusion, in general, there is not a single ellipse (among the four ellipse estimates) which represents a good segmentation for *all* the images. Let us denote by $Ell_{opt}$ the ellipse, among the four ellipse estimates ($Ell_{f1}$, $Ell_{f2}$, $Ell_{f3}$ and $Ell_{f4}$), for which $\mu_{D_{meanR}}$ is minimum. Table 2 shows the distribution of $D_{meanR}$ and $D_{maxR}$ for ellipse $Ell_{opt}$. It can be observed that the values of $D_{meanR}$ and $D_{maxR}$ for $Ell_{opt}$ are lower than for those of $Ell_{f4}$. This shows that there is at least one ellipse among the four ellipse estimates which best represents the tomato. The operator has only to select the best ellipse.
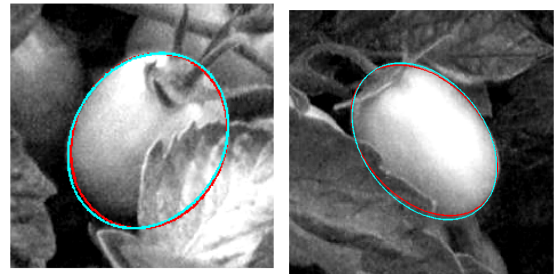
# 5 CONCLUSIONS

We presented a segmentation procedure used to monitor tomatoes in images. Starting from an approximate computation of the position of the center of the tomato, segmentation algorithms based on contour and region information are proposed and combined, in order to determine a first estimate of the contour. Then, a parametric active contour with shape constraint is applied and four ellipse estimates representing the tomatoes are obtained. In all the steps of this process, a priori knowledge about the shape and

the size of the tomatoes is modeled and incorporated as regularization terms, leading to better robustness. It is supposed that the operator selects, at the end of the process for each image, the ellipse corresponding to the best elliptic estimation of the actual contour.

Table 1: Mean ($\mu$) and standard deviation ($\sigma$) of $D_{meanR}$ and $D_{maxR}$ by comparing ellipse $Ell_{f4}$ with the manual segmentation $M$. Only the images belonging to category 1 (i.e. with a low amount of occlusion) have been considered.

| | $N_S^1$ | $\mu_{D_{meanR}}$ | $\sigma_{D_{meanR}}$ | $\mu_{D_{maxR}}$ | $\sigma_{D_{maxR}}$ |
|---|---|---|---|---|---|
| $S = 1$ | 26 | 1.72 | 0.77 | 5.06 | 2.76 |
| $S = 2$ | 4 | 1.85 | 0.46 | 5.45 | 1.91 |
| $S = 3$ | 21 | 3.40 | 2.24 | 9.79 | 6.88 |
| $S = 4$ | 14 | 2.73 | 1.92 | 7.81 | 5.71 |
| $S = 5$ | 5 | 4.81 | 1.30 | 13.05 | 3.56 |
| $S = 6$ | 0 | - | - | - | - |
| $S = 7$ | 25 | 1.88 | 0.65 | 4.81 | 1.97 |
| $S = 8$ | 20 | 6.07 | 5.75 | 15.41 | 10.61 |
| $S = 9$ | 1 | 5.26 | 0.00 | 11.86 | 0.00 |
| $S = 10$ | 5 | 2.25 | 0.56 | 6.59 | 2.25 |

Table 2: Mean ($\mu$) and standard deviation ($\sigma$) of $D_{meanR}$ and $D_{maxR}$ by comparing ellipse $Ell_{opt}$ with the manual segmentation $M$. Only the images belonging to category 1 (i.e. with a low amount of occlusion) have been considered.

| | $N_S^1$ | $\mu_{D_{meanR}}$ | $\sigma_{D_{meanR}}$ | $\mu_{D_{maxR}}$ | $\sigma_{D_{maxR}}$ |
|---|---|---|---|---|---|
| $S = 1$ | 26 | 1.34 | 0.68 | 3.76 | 2.21 |
| $S = 2$ | 4 | 1.57 | 0.4 | 4.43 | 0.88 |
| $S = 3$ | 21 | 2.87 | 2.05 | 8.4 | 6.58 |
| $S = 4$ | 14 | 2.2 | 1.77 | 6.28 | 5.34 |
| $S = 5$ | 5 | 4.54 | 1.14 | 12.44 | 3.46 |
| $S = 6$ | 0 | - | - | - | - |
| $S = 7$ | 25 | 1.7 | 0.49 | 4.61 | 1.62 |
| $S = 8$ | 20 | 5.4 | 4.88 | 14.9 | 10.37 |
| $S = 9$ | 1 | 5.24 | 0 | 11.86 | 0 |
| $S = 10$ | 5 | 1.75 | 0.36 | 4.4 | 1.25 |

The segmentation of tomatoes is a challenging task due to the presence of occlusion and variation in contrast. In order to evaluate the robustness of the proposed algorithm, the entire image set was divided into three categories based on the amount of occlusion. For the images with an acceptable level of occlusion, good results were obtained with an average variation in $D_{meanR}$ less than 6%. Also, the low standard deviation for $D_{meanR}$ indicates the robustness of the proposed algorithm. Good results with $D_{meanR} < 10\%$ were obtained on 44% of the images which contain a significant amount of occlusion.

For the moment, it has been assumed that an operator manually selects one ellipse as the final segmentation. In future work, we wish to provide automatically the best representation of the tomato. Also, in some images, the position of the tomato is not detected correctly due to the presence of other tomatoes nearby. This could be improved by updating the position of the tomato globally by considering also the movement of adjacent tomatoes. One possible improvement for the active contour model is to restrict the size of the reference ellipse, as there is little growth between two consecutive images.

## REFERENCES

Aggelopoulou, A., Bochtis, D., Fountas, S., Swain, K., Gemtos, T., and Nanos, G. (2011). Yield prediction in apple orchards based on image processing. *Journal of Precision Agriculture*, 12:448–456.

Aitkenhead, M., Dalgetty, I., Mullins, C., McDonald, A., and Strachan, N. (2003). Weed and crop discrimination using image analysis and artificial intelligence methods. *Computers and Electronics in Agriculture*, 39(3):157 – 171.

Charmi, M., Ghorbel, F., and Derrode, S. (2009). Using Fourier-based shape alignment to add geometric prior to snakes. In *ICASSP*, pages 1209–1212.

Du, C. and Sun, D. (2006). Learning techniques used in computer vision for food quality evaluation: a review. *Journal of Food Engineering*, 72(1):39 – 55.

Fischler, A. and Bolles, C. (1981). Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395.

Foulonneau, A., Charbonnier, P., and Heitz, F. (2006). Affine-invariant geometric shape priors for region-based active contours. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(8):1352–1357.

Jayas, D., Paliwal, J., and Visen, N. (2000). Review paper (automation and emerging technologies): Multilayer neural networks for image analysis of agricultural products. *Journal of Agricultural Engineering Research*, 77(2):119 – 128.

Kass, M., Witkin, A., and Terzopoulos, D. (1988). Snakes: Active contour models. *International Journal of Computer Vision*, 1(4):321–331.

Lee, W. S., Slaughter, D. C., and Giles, D. K. (1999). Robotic weed control system for tomatoes. *Precision Agriculture*, 1:95–113.

Mkhabela, M., Bullock, P., Raj, S., Wang, S., and Yang, Y. (2011). Crop yield forecasting on the Canadian prairies using MODIS NDVI data. *Agricultural and Forest Meteorology*, 151(3):385 – 393.

Narendra, V. G. and Hareesh, K. S. (2010). Prospects of computer vision automated grading and sorting systems in agricultural and food products for quality evaluation. *International Journal of Computer Applications*, 1(4):1–9.

Otsu, N. (1975). A threshold selection method from gray-level histograms. *Automatica*, 11(285-296):23–27.

Prasad, A., Chai, L., Singh, R., and Kafatos, M. (2006). Crop yield estimation model for Iowa using remote sensing and surface parameters. *International Journal of Applied Earth Observation and Geoinformation*, 8(1):26 – 33.

Stajnko, D. and Cmelik, Z. (2005). Modelling of apple fruit growth by application of image analysis. *Agriculturae Conspectus Scientificus*, 70:59–64.

Xu, C. and Prince, J. (1998). Snakes, shapes, and gradient vector flow. *IEEE Transactions on Image Processing*, 7(3):359–369.

Yang, C., Prasher, S., Landry, J., Ramaswamy, H., and Ditommaso, A. (2000). Application of artificial neural networks in image recognition and classification of crop and weeds. *Canadian Agricultural Engineering*, 42(3):147 – 152.

Zhao, H. and Pei, Z. (2013). Crop growth monitoring by integration of time series remote sensing imagery and the WOFOST model. In *2013 Second International Conference on Agro-Geoinformatics (Agro-Geoinformatics)*, pages 568–571.