Segmentation of Tomatoes in Open Field Images with Shape and Temporal Constraints

Ujjwal Verma[1]([✉]), Florence Rossant[2], Isabelle Bloch[3], Julien Orensanz[4], and Denis Boisgontier[4]

[1] Department of Electronics and Communication Engg, Manipal Institute of Technology, Manipal University, Manipal, Karnataka, India
ujjwal.verma@isep.fr, ujjwal.verma@manipal.edu
[2] Institut Superieur d'Electronique de Paris (ISEP), Paris, France
florence.rossant@isep.fr
[3] LTCI, CNRS, Télécom ParisTech, Université Paris-Saclay, Paris, France
isabelle.bloch@telecom-paristech.fr
[4] Cap2020, Gironville sur Essonne, France

Abstract. With the aim of estimating the growth of tomatoes during the agricultural season, we propose to segment tomatoes in images acquired in open field, and to derive their size from the segmentation results obtained in pairs of images acquired each day. To cope with difficult conditions such as occlusion, poor contrast and movement of tomatoes and leaves, we propose to base the segmentation of an image on the result obtained on the image of the previous day, guaranteeing temporal consistency, and to incorporate a shape constraint in the segmentation procedure, assuming that the image of a tomato is approximately an ellipse, guaranteeing spatial consistency. This is achieved with a parametric deformable model with shape constraint. Results obtained over three agricultural seasons are very good for images with limited occlusion, with an average relative distance between the automatic and manual segmentations of 6.46 % (expressed as percentage of the size of tomato).

Keywords: Image segmentation · Parametric active contours · Shape constraint · Precision agriculture

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 [10,13]. However, satellite data are affected by adverse climatic conditions (clouds, etc.) resulting in inaccurate predictions [10]. 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 [19]. 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 [4,7,11], to detect weeds in a field [2,9,18], and to model the growth of fruits and then predict the yield [1,14]. In [1], the yield of an apple orchard is estimated using only the density of flowers. In [14], only 5 images captured at different stages of the apple maturation are studied in order to predict the yield. These methods [1,14] are limited to a controlled environment (apple orchards) where complex scenarios such as occlusions are not considered. Moreover, in [1] the observed scene is modified by placing a black cloth behind the tree in order to simplify the image processing tasks. However, to the best of our knowledge, there has not been any related work where the growth of a fruit or vegetable cultivated in open fields is studied based on the images captured during the entire agricultural season.

Since there is little growth of a tomato during a given day, only one image per day is analyzed in this work, thus creating a series of 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 (Fig. 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.

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 ith 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.
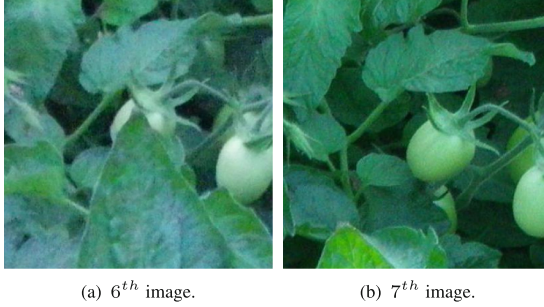
(a) 6th image.   (b) 7th image.

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

The main steps of the segmentation algorithm are as follows: 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. Thirdly, 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 Sect. 2 and the different steps of the segmentation algorithm in Sect. 3. Section 4 discusses the experimental results. A brief discussion on the second part of the system which aims at estimating the radius of the tomatoes is presented in Sect. 5. This paper extends our premilinary work in [15].

2   Active Contour with an Elliptical Shape Prior

Parametric active contour model or snake was originally introduced in [8] 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:

$$\mathbf{E}_T(v) = \mathbf{E}_{Int}(v) + \mathbf{E}_{Im}(v) + \mathbf{E}_{Ext}(v), \tag{1}$$

where $\mathbf{E}_{Int}(v)$ 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 [6] used Legendre moments to define an affine invariant shape prior in a region

based active contour. In our case, the region information is significant but not stable enough due to the presence of leaves (occlusions) and other tomatoes of similar intensity profile. In [3], 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, which is included as a constraint in 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:

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

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

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

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 [17]. The last term restricts the evolving contour to be close to the reference ellipse. The parameter α controls the smoothness of the curve, and ψ 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 θ. 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:

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

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 φ, 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 ith image. The temporal regularization (assuming little growth and movement of the tomato during a day) and the spatial regularization (tomato modeled as a sphere in the 3D space) are used throughout the segmentation procedure.

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 (Fig. 2). 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 χr^i where $r^i = \frac{a^i + b^i}{2}$ and χ is a constant determined empirically ($\chi = 1.25$ in our experiments). 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 (Fig. 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 (derived from a^i and b^i) so that we restrict the region to be analyzed as much as possible, thus reducing the computation cost of the next steps. The contrast stretching transformation is applied to imS^{i+1}.

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.

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
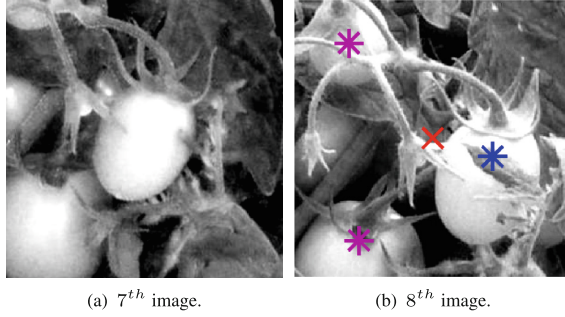
(a) 7th image.   (b) 8th image.

Fig. 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 (Color figure online).

three conditions:

$$0.5r^i < p_l < 1.5r^i \tag{5}$$

$$|\arg(\nabla imS^{i+1}(P_l)) - \theta_l| \leq \tfrac{\pi}{8} \tag{6}$$

$$|\nabla imS^{i+1}(P_l)| > \eta \tag{7}$$

where $\nabla imS^{i+1}(P_l)$ is the gradient at P_l in imS^{i+1} and η is a constant whose value is determined experimentally ($\eta = 0.2$). 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 Fig. 4(a), most points lying on the boundary of the tomato have been correctly detected along with some additional points lying on the leaves.

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 [5] estimate based on an elliptic model in order to compute several candidate ellipses. Note that the spatial (tomato modeled as an ellipse) and temporal (parameters of the model) regularization has been used in this step to increase the robustness of the segmentation procedure.

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:

$$-0.1 < \tfrac{a^{i+1}-a^i}{a^i} < 0.2, -0.1 < \tfrac{b^{i+1}-b^i}{b^i} < 0.2 \tag{8}$$

$$-0.1 < \tfrac{SA^{i+1}-SA^i}{SA^i} < 0.25 \tag{9}$$

$$|\tfrac{Ecc^{i+1}-Ecc^i}{Ecc^i}| < 0.1 \tag{10}$$

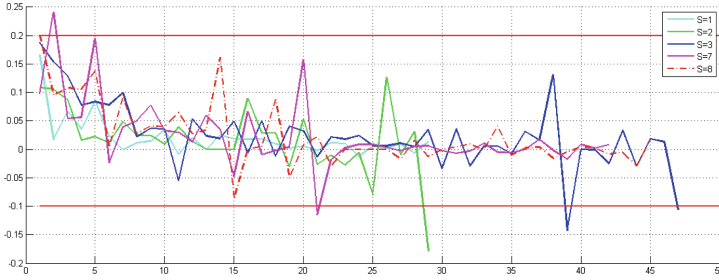$$|\tfrac{\varphi^{i+1}-\varphi^i}{\varphi^i}| < 0.2 \tag{11}$$

Fig. 3. 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 (Color figure online).

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 (Eq. 8) 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 9 restricts the apparent size of the tomato while Eq. 10 restricts the admissible values for eccentricity, thus controlling the apparent shape of the tomato.

The threshold values in Eqs. 8–11 have been determined after studying the parameters of the ellipses obtained from the manual segmentation of five tomatoes. For example, Fig. 3 shows the relative evolution of the 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 Eqs. 8–9 is due to the fact that tomatoes are supposed to grow during the agricultural season.

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 (Fig. 4(b)).

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.

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

$$\omega_{seed} = \bigcap_{u=1}^{N_a} \omega_u, \quad \omega_{limit} = \bigcup_{u=1}^{N_a} \omega_u. \tag{12}$$

The final region is denoted by ω_t (Fig. 4(c)).

We define τ_m as:

$$\tau_m = \min_{u=1,2,\dots,N_a} \tau(u), \tag{13}$$

with

$$\tau(u) = \frac{|\omega_u \cap (1 - \omega_t)| + |(1 - \omega_u) \cap \omega_t|}{|\omega_u \cap \omega_t|}, \tag{14}$$

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

Let us denote by a_u^{i+1} and b_u^{i+1} the semi-axis lengths of the candidate ellipse $Ell_u^{i+1}, u \in [1, N_a]$. We select the ellipse v (Fig. 4(d)) that minimizes $\left[\left(a_u^{i+1} - a^i\right)^2 + \left(b_u^{i+1} - b^i\right)^2\right]$ under the condition $\tau(v) \le 1.1 \; \tau_m$. 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, leading to 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.

The next step 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 ω_t and which minimizes the number of pixels inside the ellipse Ell_{te} and not belonging to the region ω_t (Fig. 4(e)). Let ω_{te} be the region inside Ell_{te}. Then, the region of occlusion ω_{oc} can be computed as $\omega_{oc} = \omega_{te} \cap \omega_t^c$.

Using morphological operations (erosion followed by reconstruction by dilation), small regions are removed from ω_{oc}, so that the resulting ω_{oc} corresponds to actual leaves causing the occlusions (Fig. 4(f)). Apart from detecting the "head" of the tomato, any other additional occlusion (mostly due to leaves) can also be detected using this approach (Fig. 4(f)).

3.5   Applying Active Contours

The active contour (Sect. 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} iterations, the parameter ψ is set to zero, so that z moves towards the most prominent contours. Then the shape constraint is introduced for $n_{ellipse}$ iterations ($\psi \ne 0$) 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 ω_{oc}, in every step of this process.
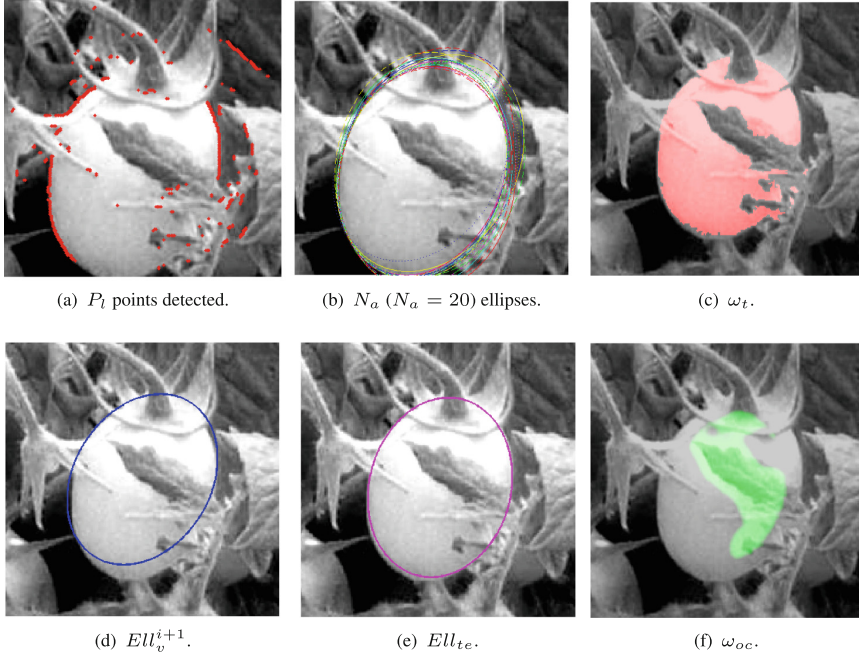
(a) P_l points detected.   (b) N_a ($N_a = 20$) ellipses.   (c) ω_t.

(d) Ell_v^{i+1}.   (e) Ell_{te}.   (f) ω_{oc}.

Fig. 4. (a,b) Points of strong gradient and ellipses detected using the RANSAC estimate. (c,d) ω_t: region representing tomato, Ell_v^{i+1}: selected initial ellipse. (e,f) Ell_{te}: convex hull of ω_t and region of potential occlusion ω_{oc}.

As explained in Sect. 2, the reference ellipse z_e is regularly 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 Sect. 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:

$$|\nabla im S^{i+1}(z(\theta)) \cdot n_e(\theta)| > \Gamma \tag{15}$$

$$\frac{|\nabla im S^{i+1}(z(\theta)) \cdot n_e(\theta)|}{|\nabla im S^{i+1}(z(\theta))|} > 0.75 \tag{16}$$

$$d(z_q(\theta), z(\theta)) < d_{max} \tag{17}$$

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 Γ is determined

(a) Updating z_e from selected points (green).

(b) Final contour z.

(c) set of points $\mathbf{P}_h$.

(d) Points $\mathbf{P}'_h$.

(e) 8th image.

(f) 17th image: $\mathbf{P}_h$ (yellow) and $\mathbf{P}'_h$ (green).
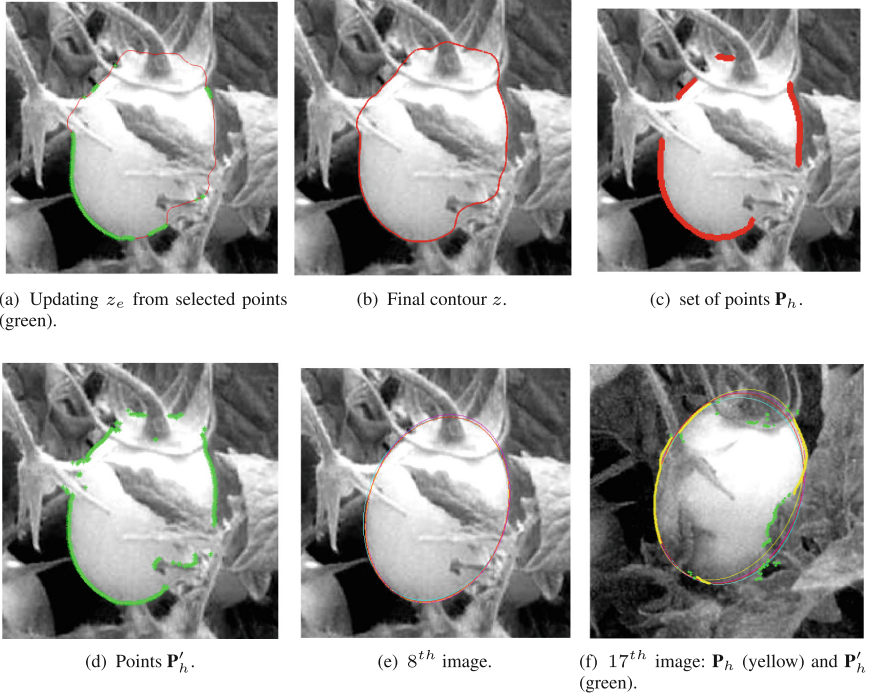
Fig. 5. (a,b) Active contour with shape constraint. (c,d) Two different sets of points $\mathbf{P}_h$ and $\mathbf{P}'_h$. (e,f) Final ellipse estimates for two different images (Color figure online).

automatically [12]. 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 considered point 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 (Fig. 5(a)).

3.6 Refining the Results

A least square estimate of an ellipse from z (Fig. 5(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 $\mathbf{P}_h$ (Fig. 5(c)) is obtained by using criteria similar to the ones described in Sect. 3.5 (Eqs. 15–17). Then, another subset $\mathbf{P}'_h$ is computed by relaxing the condition related to the gradient direction (Fig. 5(d)).

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:

$$0.9a_{f1}^{i+1} < a_{f2}^{i+1} < 1.1a_{f1}^{i+1} \tag{18}$$

$$0.9b_{f1}^{i+1} < b_{f2}^{i+1} < 1.1b_{f1}^{i+1} \tag{19}$$

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 (Fig. 5(e)). However, in case of occlusions and poor contrast, the four ellipses may be different (Fig. 5(f)), and the user selects the best one.

4   Results

Two cameras (Pentax Optio W80) were installed in an open field of tomatoes. The same setup was used for three agricultural seasons (April-August, 2011, 2012 and 2013). We have identified 21 tomatoes, covering different sites and different seasons, thus ensuring variability (614 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.

As discussed earlier, one of the main challenges of the segmentation is the occlusion and the poor quality of the images due to the poor illumination and/or shadow. Moreover, for the images acquired in the 2013 agricultural season ($S = 12, .., 21$) the size of the tomatoes was significantly smaller as compared to the one observed during the agricultural seasons in 2011 and 2012 ($S = 1, ..., 11$). This is due to the variation in the external climatic conditions. Also, in some images, a shadow created by the leaves (or the tomato itself) can be observed ($S = 8, 12$). As a result, a portion of the contour is not clear and distinct. This results in an ambiguity on the position of the contour. Given this ambiguity, even a manual segmentation is a challenging task on this portion of the contour. Moreover, a blurred contour was observed in some images of some sequences ($S = 3, 13, 18, 19, 20$), due to the presence of additional neighboring tomatoes in the background.

The data set contains images of varying contrast and degree of occlusions. Obviously, it is impossible to obtain a reliable segmentation, even manually, in case of severe occlusion. 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,
– Category 2, with $30\,\% < P < 50\,\%$ which is more prone to segmentation error,
– Category 3, with $P > 50\,\%$ for which it is impossible to perform a reliable segmentation.

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.

The obtained segmentations A were compared with the manual segmentations M (approximated by ellipses) by computing the average D^i_{mean} and maximal D^i_{max} distances between A and M for the ith 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:

$$D^i_{meanR} = \frac{D^i_{mean}}{r^i}100,\ D^i_{maxR} = \frac{D^i_{max}}{r^i}100. \tag{20}$$

For this project, $D_{meanR} < 10\,\%$ is considered as the acceptable limit of error in order to follow the growth of tomatoes.

For the images of category 1, good results (Table 1) were obtained even in the presence of occlusion by nearby leaves/branches and tomatoes (Figs. 6(a) and 6(b)). 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 (Fig. 6(c)). Also, in some images (Fig. 6(d)), due to a shadow effect on a portion of the contour, the intensity profiles of the tomato and the adjacent leaves are nearly identical, resulting in a very low contrast. Such cases may result in comparatively high distance measures even in the absence of any occlusion.

Due to the smaller size of tomatoes in sequences $S = 12, ..., 21$, higher distances were observed in these sequences (since the distances D^i_{meanR} and D^i_{maxR} are normalized by the size r^i of the tomato). For example, Fig. 6(e) shows the obtained segmentation Ell_{f4} on the 3rd image of sequence $S = 17$. The distances normalized by the size of the tomato are $D_{meanR} = 9.89\,\%$ and $D_{maxR} = 26.60\,\%$. However, the distances expressed in pixels are significantly lower ($D_{mean} = 2$, $D_{max} = 5.38$ pixels). For most of the sequences a low $\mu_{D_{meanR}}$ along with lower $\sigma_{D_{meanR}}$ demonstrates the robustness of the proposed method. However, for some sequences ($S = 13, 17, 18$) higher values of $\mu_{D_{meanR}}$ and $\sigma_{D_{meanR}}$ were observed mainly due to the false detection of the position of the tomato (Sect. 3.2), or
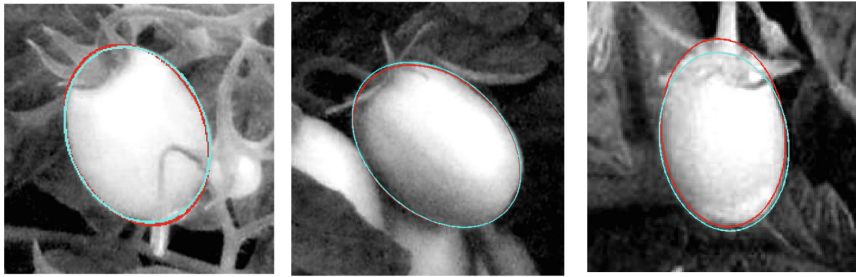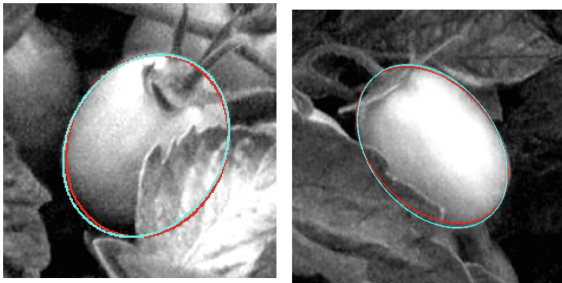
(a) Tomato $S = 1$, Image 8 (1.89%, 4.70%).
(b) Tomato $S = 4$, Image 24 (1.29%, 4.39%).
(c) Tomato $S = 5$, Image 4 (4.71%, 11.84%).
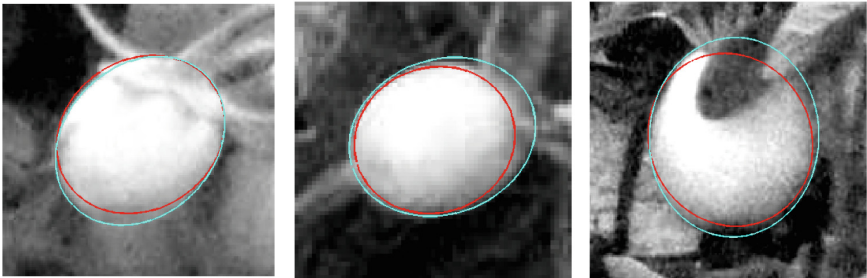
(d) Tomato $S = 8$, Image 5 (4.77%, 16.04%).
(e) Tomato $S = 17$, Image 3 (9.89%, 26.61%).
(f) Tomato $S = 12$, Image 7 (8.13%, 19.80%).

Fig. 6. Final segmentation Ell_{f4} (red) obtained on images of category 1, and values of (D_{meanR}, D_{maxR}) with respect to the manual segmentation (in cyan) (Color figure online).

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

Fig. 7. Final segmentation Ell_{f4} (red) obtained on images of category 2 (Color figure online).

due to the small size of the tomato, as discussed previously. For the sequence $S = 11$, all the images suffer from poor contrast and noise due to the shadow created by leaves. As a result, even a manual segmentation is a challenging task in this sequence.

Table 1. Mean (μ) and standard deviation (σ) of D_{meanR} and D_{maxR} by comparing ellipse Ell_{f4} and 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.

		Ell_{f4}				Ell_{opt}			
	N_S^1	$\mu_{D_{meanR}}$	$\sigma_{D_{meanR}}$	$\mu_{D_{maxR}}$	$\sigma_{D_{maxR}}$	$\mu_{D_{meanR}}$	$\sigma_{D_{meanR}}$	$\mu_{D_{maxR}}$	$\sigma_{D_{maxR}}$
$S = 1$	26	1.72	0.77	5.06	2.76	1.34	0.68	3.76	2.21
$S = 2$	4	1.85	0.46	5.45	1.91	1.57	0.40	4.43	0.88
$S = 3$	21	3.4	2.24	9.79	6.88	2.87	2.05	8.40	6.58
$S = 4$	14	2.73	1.92	7.81	5.71	2.20	1.77	6.28	5.34
$S = 5$	5	4.81	1.3	13.05	3.56	4.54	1.14	12.44	3.46
$S = 6$	0	-	-	-	-	-	-	-	-
$S = 7$	25	1.88	0.65	4.81	1.97	1.7	0.49	4.61	1.62
$S = 8$	20	6.07	5.75	15.41	10.61	5.4	4.88	14.9	10.37
$S = 9$	1	5.26	0	11.86	0	5.24	0.00	11.86	0.00
$S = 10$	5	2.25	0.56	6.59	2.25	1.75	0.36	4.40	1.25
$S = 11$	4	11.81	4.99	32.66	11.39	10.18	5.38	29.21	14.74
$S = 12$	19	4.74	1.33	11.69	3.23	4.25	1.34	11.1	3.44
$S = 13$	5	41.5	16.55	84.98	27.57	40.48	16.18	83.51	26.85
$S = 14$	4	9.57	2.35	29.48	7.3	9.18	2.57	28.06	7.84
$S = 15$	0	-	-	-	-	-	-	-	-
$S = 16$	21	4.68	1.08	10.22	2.74	4.46	1.18	10.12	2.68
$S = 17$	20	11.78	2.44	26.75	4.21	11.56	2.48	26.85	4.04
$S = 18$	23	14.18	20.06	35.76	33.83	13.94	20.09	35.15	33.42
$S = 19$	0	-	-	-	-	-	-	-	-
$S = 20$	5	8.76	5.38	20.05	14.93	6.88	2.3	13.65	2.98
$S = 21$	25	7.34	3.18	16.56	8.93	7.12	3.15	15.76	8.94

For the images of category 2 containing a significant amount of occlusion, D_{meanR} is significantly higher than for images of category 1. This is because of heavy occlusions along with the poor quality of images (effects due to shadow and/or presence of other tomatoes). However, in some sequences ($S = 1, 2, 5, 6, 8, 9, 15, 18, 20, 21$) an average D_{meanR} of less than 10% was observed. This is because of the good contrast on the non-occluded arc in these images, which results in a good segmentation. Finally, good results were obtained on 73% of the images (Fig. 7), where $D_{meanR} < 10\%$ even in the presence of severe occlusions.

In the results presented so far, Ell_{f4} was compared with the manual segmentation. However Ell_{f4} is not necessarily the best ellipse, and was selected here for illustrative purpose only. Due to the variation in the contrast and occlusion, 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 1 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}. The operator selects Ell_{opt} as the final segmentation.

5   Estimating the Size of the Tomato

From the obtained segmentation in both images and the camera parameters, we then estimate the size of the tomatoes. However, determining the image point pairs which correspond to the same point in the 3D space is a challenging task given the complexity of the scene. Instead we simplify the size estimation procedure by exploiting the spherical hypothesis.

The contour of the tomato is approximated by an ellipse whose parameters are calculated using the procedure presented above. Then, the sphere center in the 3D space is computed using triangulation from the centers of the ellipse calculated in both images. Next, the 3D space points situated on the contour of the tomato are computed using properties of projective geometry, independently from each image. Finally, a joint optimization procedure enables us to estimate the sphere radius.

In order to evaluate the size estimation procedure, the size of tomatoes observed in laboratory was measured. Since a tomato is not a perfect sphere, two reference values were measured manually and compared with the estimated radius of the sphere. For the manually segmented tomatoes observed in laboratory, we found that the relative percentage error between the largest of the reference value and the estimated radius was less than 5 % in 91 % of the cases. For the tomatoes cultivated in the open field, the relative percentage error was less than 10 % in 80 % of the cases [16]. The errors are mainly caused by the imperfect segmentation, due to shadowing effect and the poor quality of the images.

6   Conclusions

We presented a segmentation procedure used to monitor the growth of tomatoes from images acquired in an open field. 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 a 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.

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 on most (87 %) of images where D_{meanR} was less than 10 %. Also, the low standard deviation for D_{meanR} indicates the robustness of the proposed algorithm. Good results with $D_{meanR} < 10$ % were obtained on 73 % of the images that 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

1. Aggelopoulou, A., Bochtis, D., Fountas, S., Swain, K., Gemtos, T., Nanos, G.: Yield prediction in apple orchards based on image processing. J. Precis. Agric. **12**, 448–456 (2011)
2. Aitkenhead, M., Dalgetty, I., Mullins, C., McDonald, A., Strachan, N.: Weed and crop discrimination using image analysis and artificial intelligence methods. Comput. Electron. Agric. **39**(3), 157–171 (2003)
3. Charmi, M., Ghorbel, F., Derrode, S.: Using Fourier-based shape alignment to add geometric prior to snakes. In: ICASSP, pp. 1209–1212 (2009)
4. Du, C., Sun, D.: Learning techniques used in computer vision for food quality evaluation: a review. J. Food Eng. **72**(1), 39–55 (2006)
5. Fischler, A., Bolles, C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun. ACM **24**(6), 381–395 (1981)
6. Foulonneau, A., Charbonnier, P., Heitz, F.: Affine-invariant geometric shape priors for region-based active contours. IEEE Trans. Pattern Anal. Mach. Intell. **28**(8), 1352–1357 (2006)
7. Jayas, D., Paliwal, J., Visen, N.: Review paper (automation and emerging technologies): multi-layer neural networks for image analysis of agricultural products. J. Agric. Eng. Res. **77**(2), 119–128 (2000)
8. Kass, M., Witkin, A., Terzopoulos, D.: Snakes: active contour models. Int. J. Comput. Vis. **1**(4), 321–331 (1988)
9. Lee, W.S., Slaughter, D.C., Giles, D.K.: Robotic weed control system for tomatoes. Precis. Agric. **1**, 95–113 (1999)
10. Mkhabela, M., Bullock, P., Raj, S., Wang, S., Yang, Y.: Crop yield forecasting on the Canadian prairies using MODIS NDVI data. Agric. Forest Meteorol. **151**(3), 385–393 (2011)
11. Narendra, V.G., Hareesh, K.S.: Prospects of computer vision automated grading and sorting systems in agricultural and food products for quality evaluation. Int. J. Comput. Appl. **1**(4), 1–9 (2010)
12. Otsu, N.: A threshold selection method from gray-level histograms. Automatica **11**(285–296), 23–27 (1975)

13. Prasad, A., Chai, L., Singh, R., Kafatos, M.: Crop yield estimation model for Iowa using remote sensing and surface parameters. Int. J. Appl. Earth Obs. Geoinf. **8**(1), 26–33 (2006)
14. Stajnko, D., Cmelik, Z.: Modelling of apple fruit growth by application of image analysis. Agric. Conspec. Sci. **70**, 59–64 (2005)
15. Verma, U., Rossant, F., Bloch, I., Orensanz, J., Boisgontier, D.: Shape-based segmentation of tomatoes for agriculture monitoring. In: International Conference on Pattern Recognition Applications and Methods (ICPRAM), Angers, France, pp. 402–411 (2014)
16. Verma, U., Rossant, F., Bloch, I., Orensanz, J., Boisgontier, D.: Tomato development monitoring in an open field, using a two-camera acquisition system. In: 12th International Conference on Precision Agriculture (2014)
17. Xu, C., Prince, J.: Snakes, shapes, and gradient vector flow. IEEE Trans. Image Process. **7**(3), 359–369 (1998)
18. Yang, C., Prasher, S., Landry, J., Ramaswamy, H., Ditommaso, A.: Application of artificial neural networks in image recognition and classification of crop and weeds. Can. Agric. Eng. **42**(3), 147–152 (2000)
19. Zhao, H., Pei, Z.: 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), pp. 568–571 (2013)