



# Segmentation and size estimation of tomatoes from sequences of paired images

Ujjwal Verma, Florence Rossant, Isabelle Bloch

## ► To cite this version:

Ujjwal Verma, Florence Rossant, Isabelle Bloch. Segmentation and size estimation of tomatoes from sequences of paired images. EURASIP Journal on Image and Video Processing, 2015, 10.1186/s13640-015-0087-0 . hal-01228926

**HAL Id: hal-01228926**

**<https://hal.science/hal-01228926>**

Submitted on 15 Dec 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

RESEARCH

Open Access



# Segmentation and size estimation of tomatoes from sequences of paired images

Ujjwal Verma<sup>1\*</sup>, Florence Rossant<sup>2</sup> and Isabelle Bloch<sup>3</sup>

## Abstract

In this paper, we present a complete system to monitor the growth of tomatoes from images acquired in open fields. This is a challenging task because of the severe occlusion and poor contrast in the images. We approximate the tomatoes by spheres in the 3D space, hence by ellipses in the image space. The tomatoes are first identified in the images using a segmentation procedure. Then, the size of the tomatoes is measured from the obtained segmentation and camera parameters. The shape information combined with temporal information, given the limited evolution from an image to the next one, is used throughout the system to increase the robustness with respect to occlusion and poor contrast.

The segmentation procedure presented in this paper is an extension of our previous work based on active contours. Here, we present a method to update the position of the tomato by comparing the SIFT descriptors computed at predetermined points in two consecutive images. This leads to a very accurate estimation of the tomato position, from which the entire segmentation procedure benefits. The average error between the automatic and manual segmentations is around 4 % (expressed as the percentage of tomato size) with a good robustness with respect to occlusion (up to 50 %).

The size estimation procedure was evaluated by calculating the size of tomatoes under a controlled environment. In this case, the mean percentage error between the actual radius and the estimated size is around 2.35 % with a standard deviation of 1.83 % and is less than 5 % in most (91 %) cases. The complete system was also applied to estimate the size of tomatoes cultivated in open fields.

**Keywords:** Image segmentation, Parametric active contours, Shape constraint, Precision farming, Metric measurement

## 1 Introduction

Monitoring the growth of crop provides important information about the status of the crop and helps the farmer in better managing resource requirements (such as storage requirements, transportation) after the harvest. It also allows better planning and marketing well in advance, as well as better negotiating terms and condition for crop insurance. Moreover, any abnormal growth of the crop can be determined through crop continuous monitoring during the entire agriculture season [1].

Existing methods for monitoring the growth of crops can be broadly divided into two categories. In the first category, the growth of the crop is monitored and the yield of

the field is estimated based on remote sensing data [2, 3]. Various vegetation indices such as normalized difference vegetation index (NDVI) and vegetation condition index (VCI) are used to calculate the growth stage of the crop and then estimate the yield. However, the quality of the acquired data may decrease due to adverse climatic condition (such as clouds) [3]. Moreover, since NDVI is based on reflected radiation in the near-infrared and visible wavelengths, the condition of the soil could result in unreliable measured indices. Crop growth modeling is another method used to model the growth of the crop based on crop variety, soil, and weather information [4]. The drawback of this method is that it considers an ideal scenario with no infection in the field. In case of an infection, the estimated model would not accurately represent the actual growth status of the crop.

\*Correspondence: [ujjwal.verma@manipal.edu](mailto:ujjwal.verma@manipal.edu)

<sup>1</sup>Department of Electronics and Communication Engineering, Manipal Institute of Technology, Manipal University, Manipal, India  
Full list of author information is available at the end of the article

There exist few studies where the growth of the crops is monitored based on captured images of the field [5, 6]. For instance, the authors in [5] proposed to detect apples cultivated in apple orchards and then measure their size based on morphological operations. However, the proposed method does not take into account any possible occlusion, which may be a strong limitation. The authors in [6] developed a model that can predict the yield of the field at harvest, given the flower density calculated from the captured image. In order to have maximal contrast in the image and least influence of sunlight condition, a black screen made of textile was placed behind the trees, which is a heavy and painful task. Moreover, the yield of the field at harvest depends not only on the flower density but also on the meteorological conditions during the season. These methods are limited to a controlled environment where there is little occlusion and little movement between consecutive images.

This study proposes an innovative system for monitoring the growth of tomatoes cultivated in open fields, from images acquired at regular time intervals by two cameras. The purpose is to estimate the size of the tomatoes remotely, all along their maturation (from flowers to ripe fruits), in order to detect any abnormal development and predict the yield of the field. No specific installation is required to control the environment. The two cameras are placed in the open field and the images are transferred to a central server via a wireless network (2G, 3G, 4G). The data are stored and analyzed at the central server. Note that a judicious amount of data should be transferred to the central server in order to minimize the cost. From the estimated sizes, any abnormal development can be deduced and an estimate of the yield of the field can be computed remotely.

The growth monitoring and yield estimation require identifying the tomatoes in the images and performing quantitative measurements based on calibrated acquisitions. Figure 1 shows typical images of the field using a two-camera acquisition system. Given the challenges of

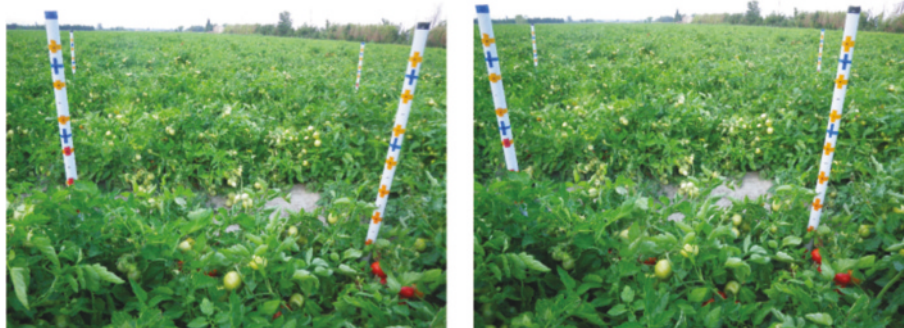
the system (see Section 2), we suppose that a tomato can be modeled as a sphere in the 3D space. Therefore, the tomatoes are first identified in the images using a segmentation procedure, which exploits shape information combined with temporal information, given the limited evolution from an image to the next one. The size of the tomatoes is then measured from the obtained segmentation and camera parameters. We chose to capture and analyze images of the tomatoes during their whole development while keeping costs as low as possible. We chose to process one pair of images per day, as analyzing images more frequently would increase the cost of transmission and processing, without bringing significant extra information.

This work is an extension of our previous works [7, 8] which presented the segmentation procedure only. In contrast, this paper presents the complete system to follow the growth of tomatoes, which consists in segmenting the tomatoes and then estimating their sizes. Furthermore, we propose a more accurate method for estimating the position of the tomato as compared to our previous works [7, 8], thereby improving the obtained segmentation and leading to more accurate experimental results.

As it can be observed from Fig. 1, detecting the tomatoes is a very challenging task. This point is discussed in Section 2, and the model we propose to overcome the difficulties is presented in Section 3. Sections 4 and 5 describe the proposed methods for the two parts of the system: segmentation procedure and size estimation procedure. Experimental results obtained with our complete system on data acquired in open fields are presented in Section 6.

## 2 Challenges of the system

One of the major challenges of the system results from occlusion. Most of the tomatoes are either partially or completely hidden by other tomatoes or branches/leaves. Figure 2 shows images of three different tomatoes. It is worth noting the variation in the amount of occlusion. In



**Fig. 1** Images of the scene acquired using two cameras



**Fig. 2** Variation in the amount of occlusion

some cases, due to severe occlusion, it is impossible to correctly detect the tomatoes even manually.

Color information is not very useful as the tomatoes and leaves are almost of the same color during a major part of the agriculture season (Fig. 1). Moreover, the position of the tomato is not fixed during an agriculture season. This might be due to external climatic condition (wind, rain) or due to the increasing weight of the tomato as the season progresses.

Since the images are acquired in open fields, we do not have any control on the external illumination of the scene. As a result, a shadow is observed in some images (Fig. 3). Moreover, due to the presence of neighboring tomatoes in the background, a portion of the contour is blurred (Fig. 3). This results in an imprecision regarding the delineation of the actual position of the contour.

In order to measure the size of the tomato, we need to determine the image points in the two images corresponding to the same point in the 3D space. This correspondence problem is another very challenging task given the complexity of the scene.

In order to overcome all these problems, we propose to exploit available a priori information. The next section introduces our system and describes how this information is integrated.

### 3 Proposed system

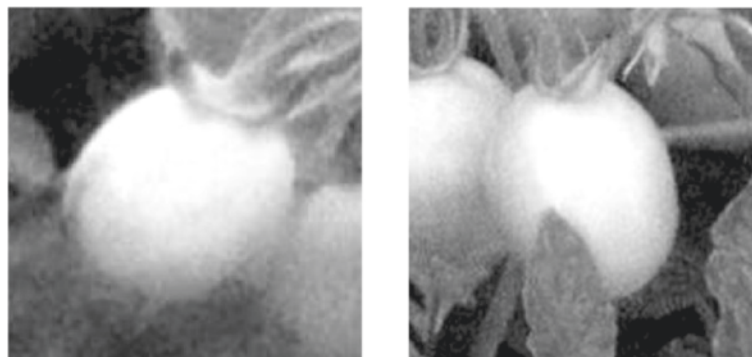
We suppose that the tomato is a sphere in the 3D space. Using the properties of projective geometry, it can be shown that the image of a sphere in the 3D space is an ellipse (Section 3.1). Moreover, we found experimentally that the ellipse parameters vary slowly from one day to the next one, allowing us to introduce temporal knowledge in our models (Section 3.2). All this information is used all along our segmentation method which relies on an active contour model with shape constraint (Section 3.3). The complete workflow is introduced in Section 3.4.

Figure 4 shows the geometry of the acquisition system. The camera parameters (intrinsic and extrinsic) were computed by observing a calibration pattern at different locations and orientations in the scene [9].

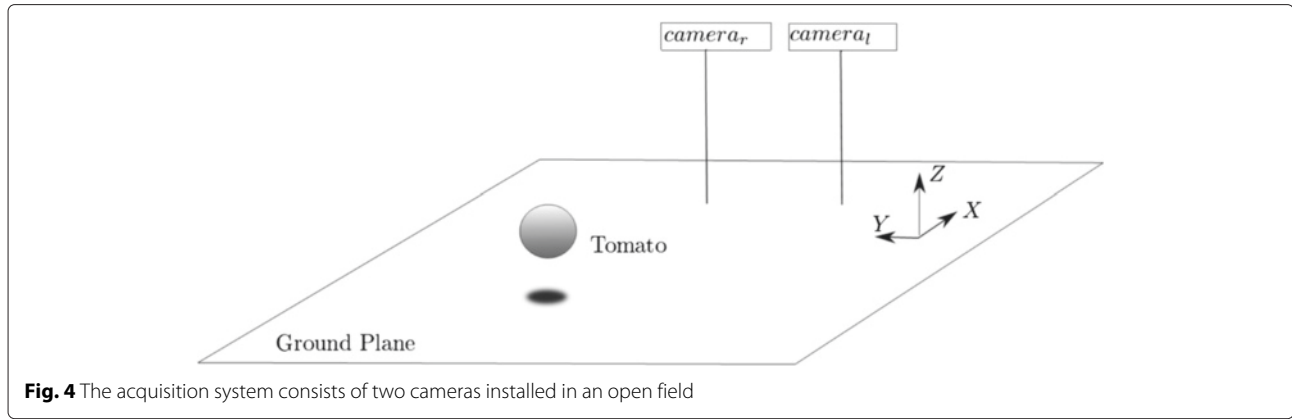
#### 3.1 Geometric model

The contour generator  $\Gamma$  of a surface  $\mathbf{Qr}$  of the 3D space, in general, is a space curve, composed of all points  $\mathbf{X}$  situated on the surface at which the imaging rays are tangent. The apparent contour  $\mathbf{Cn}$  is the image of this contour generator.

Under the camera projection matrix  $\mathbf{P}$ , the apparent contour of the quadric  $\mathbf{Qr}$  (a surface in the 3D space) is the conic  $\mathbf{Cn}$  defined as:



**Fig. 3** Imprecision on the actual position of the contour due to shadow (left) and blurring (right)



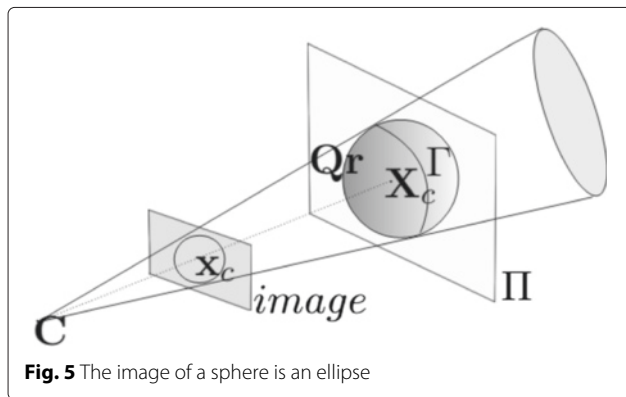
**Fig. 4** The acquisition system consists of two cameras installed in an open field

$$\mathbf{Cn}^* = \mathbf{PQr}^*\mathbf{P}^T \quad (1)$$

where  $\mathbf{M}^*$  represents the adjoint of  $\mathbf{M}$  or  $\mathbf{M}^* \propto \mathbf{M}^{-1}$  for a non-singular matrix  $\mathbf{M}$ , where  $\propto$  denotes equality up to a scale factor. Note that a dual conic  $\mathbf{Cn}^*$  is used here because apparent contour arises from tangency (see Chapter 8 in [10]).

In our case, the contour generator  $\Gamma$  of a sphere is a circle whose center is the center  $\mathbf{X}_c$  of the sphere (Fig. 5). The curve  $\Gamma$  is included in the plane  $\Pi$  orthogonal to the line joining the center of the sphere  $\mathbf{X}_c$  and the camera center  $\mathbf{C}$ . The rays from the camera center  $\mathbf{C}$  and tangent at the contour generator  $\Gamma$  form a cone, with the camera center as its vertex. The intersection of the image plane with the cone gives the image of the contour generator. Therefore, the apparent contour of this contour generator (circle) is an ellipse.

Introducing this a priori shape information in the segmentation procedure increases the robustness of the segmentation with respect to occlusion. Besides, this also simplifies the size estimation procedure since the radius of the sphere can be estimated without a full 3D reconstruction of the scene.



**Fig. 5** The image of a sphere is an ellipse

### 3.2 Temporal model

As discussed earlier, there is little growth of the tomato in a given day. Therefore, only two images are acquired every day, one for each camera. This creates a sequence of images for a given tomato. We manually segmented five tomatoes and approximated the delineated contour by an ellipse. We then studied the evolution of the length of their major and minor axes for the entire agriculture season. This study confirmed that there is little growth in the tomato during a given day [7].

Moreover, it was observed that under normal circumstances, there is a little movement of the tomato as the season progresses. However, this movement is not uniform and very difficult to predict especially in case of strong winds or heavy rains, which led us to propose a new algorithm for tomato detection, presented in Section 4.2.

This a priori temporal information is integrated in the segmentation algorithm by defining intervals which represent the acceptable variation of the ellipse parameters in a given day. The admissible range of variation of the ellipse parameters is quite wide and defined as follows:

$$-0.1 < \frac{a^{i+1} - a^i}{a^i} < 0.2 \quad (2)$$

$$-0.1 < \frac{b^{i+1} - b^i}{b^i} < 0.2 \quad (3)$$

$$-0.1 < \frac{SA^{i+1} - SA^i}{SA^i} < 0.25 \quad (4)$$

$$\left| \frac{Ecc^{i+1} - Ecc^i}{Ecc^i} \right| < 0.1 \quad (5)$$

$$\left| \frac{\varphi^{i+1} - \varphi^i}{\varphi^i} \right| < 0.2 \quad (6)$$

where  $a^i, b^i, \varphi^i, SA^i$ , and  $Ecc^i$  are the semi major axis length, semi minor axis length, orientation, area, and eccentricity of the ellipse in the previous image ( $i^{\text{th}}$  image), respectively.



### 3.3 Active contour with shape constraint

The proposed segmentation procedure is based on an active contour model [11] with shape constraint. A brief description of the proposed active contour model [7] is presented below.

As discussed, the image of the tomato is assumed to be an ellipse. Thus, we propose to use a parametric active contour with an elliptic shape constraint, derived from a reference ellipse. Assuming the center of the reference ellipse as the origin, a point  $z(\theta)$  on the evolving contour can be defined as:

$$z(\theta) = r(\theta)e^{j\theta} \quad (7)$$

where  $\theta$  is the angular coordinate and  $r(\theta)$  is the radial coordinate. Similarly, a point situated on the reference ellipse can be denoted as  $z_e(\theta) = r_e(\theta)e^{j\theta}$ . The energy functional with shape prior is defined as:

$$E_{\text{Total}}(r, r_e) = \int_0^{2\pi} \left( \frac{\alpha}{2} |r'(\theta)|^2 \right) d\theta + \int_0^{2\pi} E_{\text{Image}}(r(\theta)e^{j\theta}) d\theta + \frac{\psi}{2} \int_0^{2\pi} |r(\theta) - r_e(\theta)|^2 d\theta \quad (8)$$

In the above equation, the internal energy is represented by the first term and the external energy by the last term, which restricts the evolution of the contour with respect to the reference ellipse. The coefficient  $\alpha$  controls the variations of  $r$  and makes it regular, while  $\psi$  controls the influence of the shape prior on the total energy. In our application, these two parameters were set experimentally ( $\alpha = 10$ ,  $\psi = 0.5$ ) on some examples and the same values were used for all images. The second term, the image energy term, is calculated using gradient vector flow (GVF) [12]. The minimization of Eq. 8 is classically performed thanks to an iterative algorithm. The reference ellipse is regularly updated, based on both the position of the current curve  $z(\theta)$  and the knowledge of the final ellipse in the previous image (temporal model).

### 3.4 Summary of the proposed algorithm

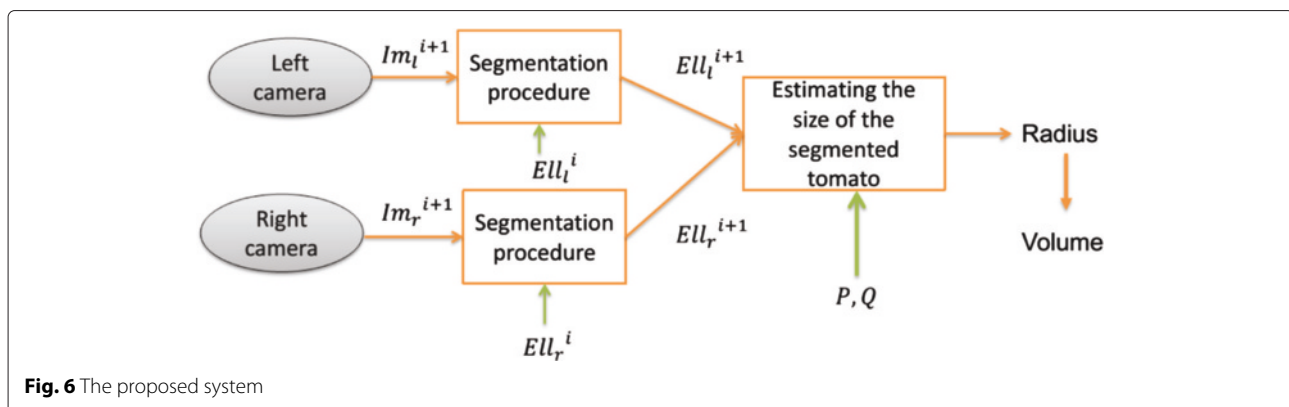
Here, we briefly present the main steps for processing a tomato in a series of images captured by the left and right cameras. The global system is composed of two main steps: tomato segmentation and size estimation (Fig. 6). We assume that the images of a given day ( $i+1$ ),  $\text{Im}_l^{i+1}$  and  $\text{Im}_r^{i+1}$ , are processed knowing the elliptic approximations,  $\text{Ell}_l^i$  and  $\text{Ell}_r^i$ , found in the previous day ( $i$ ).

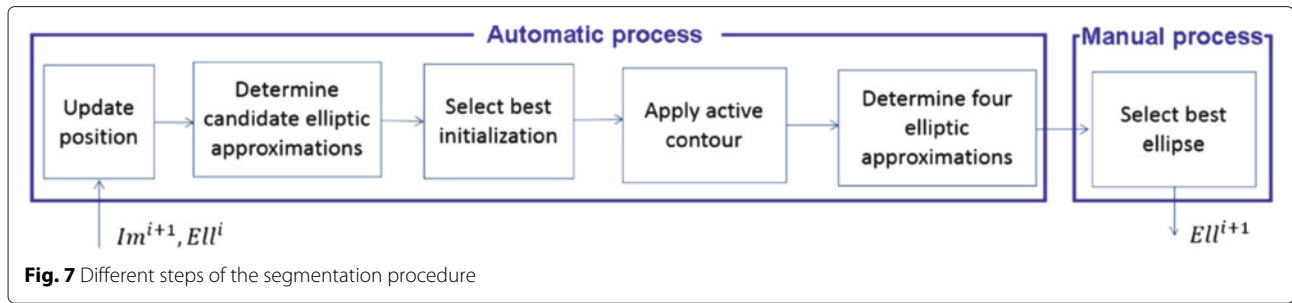
We first summarize the segmentation procedure applied to the left and right images separately (Fig. 7). Since the position of the tomato is not fixed in the image as the season progresses, we first update it (Section 4.2). Then, gradient information is combined with region information in order to propose a first elliptic approximation of the tomato boundary, from which the active contour with elliptic shape constraint can be applied. Finally, four ellipse estimates are computed from which the operator has to select the best one as the final segmentation (Section 4.3).

In order to perform metric measurements in the scene, we need to determine the camera projection matrices for the two cameras ( $\mathbf{P}, \mathbf{Q}$ ). These matrices are computed using the method presented in [9]. Then, from the obtained segmentation in the two images ( $\text{Ell}_l^{i+1}, \text{Ell}_r^{i+1}$ ) and the camera projection matrices, the set of 3D space points situated on the contour generator is computed (Section 5.1). From the two sets of 3D space points corresponding to the left and the right cameras, we then estimate the radius of the sphere using least square minimization techniques (Section 5.2). Finally, a joint optimization is performed to obtain the final radius estimate of the sphere (Section 5.3).

## 4 Segmentation procedure

This section presents the proposed algorithm for detecting the tomatoes. Let us denote by  $\text{Im}^{i+1}$  the  $(i+1)^{\text{th}}$  image (left or right) in which we wish to identify the tomato. In our sequential approach, the contour in the  $(i+1)^{\text{th}}$  image is computed based on the information present in  $\text{Im}^{i+1}$  and the contour of the tomato in the





$i^{\text{th}}$  image ( $Im^i$ ) which has been validated by the operator (Fig. 7). So, in the following steps, it is assumed that the contour representing the tomato in the  $i^{\text{th}}$  image is available and reliable. It is denoted by the ellipse  $Ell_f^i = [xc_f^i, yc_f^i, a_f^i, b_f^i, \varphi_f^i]$  where  $C_f^i = [xc_f^i, yc_f^i]$  represents the center of the ellipse whose semi major and minor axes lengths are  $a_f^i$  and  $b_f^i$ , respectively, and which has a rotation angle of  $\varphi_f^i$ .

#### 4.1 Pre-processing

Color information is not very useful since the tomatoes turn to red only at the end of the season. However, the edges of the tomatoes are more contrasted in the red component of the image, even during the first stages of the maturation. Hence, only this component is considered. A contrast stretching transformation is applied to this image.

#### 4.2 Tomato localization

We first update the position of the tomato in the current image ( $(i+1)^{\text{th}}$ ) using a descriptor-based approach. Given the complexity of the scene, detecting interest points in the entire image and then matching their descriptors would be a computationally expensive task. Instead, we propose to compare the descriptors computed at a pre-determined set of points in the previous ( $i^{\text{th}}$ ) and current ( $(i+1)^{\text{th}}$ ) images. The points in the  $i^{\text{th}}$  image are computed from the final segmentation validated by the operator (Section 4.3.3). As such, these points are very likely to be situated on the actual boundary of the tomato, or they are very close to it otherwise. In the  $(i+1)^{\text{th}}$  image, the candidate points are computed based on gradient magnitude and direction; they may or may not lie on the actual boundary of the tomato. By matching these two sets of descriptors, we then compute the translation undergone by the tomato from the  $i^{\text{th}}$  image to the  $(i+1)^{\text{th}}$  image.

##### 4.2.1 Selection of relevant points in the $i^{\text{th}}$ image

Considering the final segmentation  $v_f^i$  (output by the active contour algorithm) and its least square estimate  $Ell_f^i$  in the  $i^{\text{th}}$  image (Fig. 8), the first step aims at selecting from  $v_f^i$  a set  $P^i$  of points that are surely situated on the tomato.

For each candidate point  $P_{v,h}^i$  of  $v_f^i$ ,  $h = 1, \dots, n_{P_v}^i$ , the nearest point  $Q_{v,h}^i$  situated on the ellipse  $Ell_f^i$  is first determined. The normal to the ellipse  $Ell_f^i$  at the point  $Q_{v,h}^i$  is calculated and denoted by  $n_h^i$ .

In order to search out the points in the neighborhood of  $P_{v,h}^i$  with prominent gradient magnitude and whose gradient direction is normal to the ellipse  $Ell_f^i$ , an intensity profile is created between  $P_{h,1}^i$  and  $P_{h,2}^i$ , where  $P_{h,1}^i = P_{v,h}^i - 0.25 r_f^i n_h^i$  and  $P_{h,2}^i = P_{v,h}^i + 0.1 r_f^i n_h^i$ , with

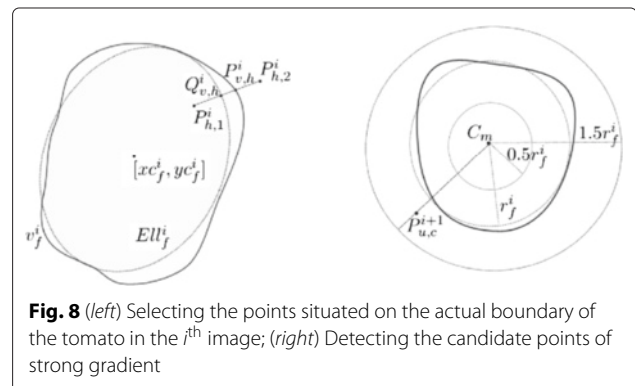
$$r_f^i = \frac{a_f^i + b_f^i}{2} \quad (9)$$

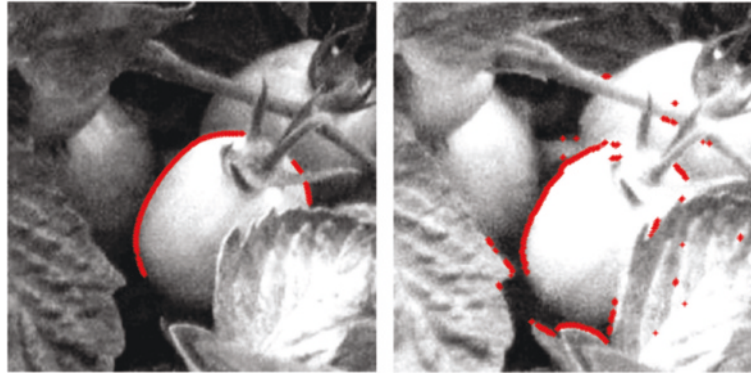
If we denote by  $P_{h,\max}^i$  the point with maximum gradient magnitude along the segment  $[P_{h,1}^i, P_{h,2}^i]$ , then  $P_{v,h}^i$  is selected as a point situated on the actual contour of the tomato if:

$$d(P_{h,\max}^i, P_{v,h}^i) < 2 \quad (10)$$

where  $d$  represents the Euclidean distance. Hereinafter, the set of points situated on the actual contour of the tomato computed using the above condition is represented as  $P^i = \{P_h^i, h = 1, 2, \dots, n_P^i\}$ , where  $n_P^i \leq n_{P_v}^i$ .

Figure 9 shows the points detected in the 17th image of sequence  $S = 7$ . Most of the non-occluded contour points of the tomato have been selected using this approach.





**Fig. 9** The two sets of points to be matched: (left)  $P^i$  in  $S = 7, i = 17$ , (right)  $P_c^{i+1}$  in  $S = 7, i = 18$

#### 4.2.2 Selection of candidate points in the $(i + 1)^{\text{th}}$ image

We now wish to compute a set of candidate contour points in the  $(i + 1)^{\text{th}}$  image. We first roughly determine the position of the tomato in the  $(i + 1)^{\text{th}}$  image, based on the pattern matching method presented in [7]. Let us denote by  $C_m = [x_m, y_m]$  the estimated position. The purpose of the proposed method is to refine this position.

Using a polar representation with  $C_m$  as the origin, let us denote by  $\{P_{u,c}^{i+1} = \rho_{u,c}^{i+1} e^{j\theta_{u,c}^{i+1}}, c = 1, \dots, n_{P_u}^{i+1}\}$  the set of points situated inside two concentric circles (Fig. 8) whose radii are respectively  $0.5r_f^i$  and  $1.5r_f^i$ . Points from this set are selected as candidate points if they satisfy the following two conditions:

$$|\arg(\nabla \text{Im}^{i+1}(P_{u,c}^{i+1})) - \theta_{u,c}^{i+1}| \leq \delta\theta_{\max} \quad (11)$$

$$|\nabla \text{Im}^{i+1}(P_{u,c}^{i+1})| > \eta \quad (12)$$

where  $\arg(\nabla \text{Im}^{i+1}(P_{u,c}^{i+1}))$  and  $|\nabla \text{Im}^{i+1}(P_{u,c}^{i+1})|$  are respectively the angle and magnitude of the gradient at  $P_{u,c}^{i+1}$  in  $\text{Im}^{i+1}$ .

The threshold values have been determined experimentally ( $\eta = 0.2$  pixels,  $\delta\theta_{\max} = \frac{\pi}{8} \text{ rad}$ ). The above conditions can be viewed as selecting the points with strong gradient whose gradient direction is within an acceptable limit with respect to the normal vector to a circle with radius  $r_f^i$ . Finally, for every angle  $\theta_{u,c}^{i+1}$ , if several points satisfy the criteria defined in Eqs. 11 and 12, only the closest to the center point  $C_m$  is retained. Thus, at most, one candidate point is retained by angle. This allows reducing the number of candidate points to be processed, in a way that is consistent with the fact that the non-occluded pixels of the tomato do not have prominent gradient.

Hereinafter, the set of candidate points in the  $(i + 1)^{\text{th}}$  image is denoted by  $P_c^{i+1} = \{P_{c,l}^{i+1}, l = 1, \dots, n_P^{i+1}\}$ ,  $n_P^{i+1} \leq n_{P_u}^{i+1}$ .

Figure 9 shows the set of points  $P_c^{i+1}$  detected for the 18th image of sequence  $S = 7$ . Note that the points lying on the actual contour of the tomato have been detected along with several other points lying on the adjacent leaves.

#### 4.2.3 Descriptor matching

Next, we wish to match the descriptors computed at these two sets of points. Let us denote by  $D_{P^i}$  and  $D_{P_c^{i+1}}$  the scale invariant feature transform (SIFT) descriptors [13] computed at  $P^i$  and  $P_c^{i+1}$ , respectively. For every interest point of  $P^i$ , the best match is determined by minimizing the Euclidean distance between the two descriptor vectors (using the keypoint matching approach presented in [13]). Let us suppose that the  $k^{\text{th}}$  point of the set  $P^i$  matches with the  $l^{\text{th}}$  point in the set  $P_c^{i+1}$ . The corresponding translation  $T_k^i$  is then computed as:

$$T_k^i = P_{c,l}^{i+1} - P_k^i, k = 1, \dots, n_P^i \quad (13)$$

Now, we wish to determine which translation among the  $n_P^i$  possible candidates represents the actual movement of the tomato. For each candidate, we define the translated set of points  $P_{T_k}^i = \{P_{T_{k,h}}^i\}$  where

$$P_{T_{k,h}}^i = P_h^i + T_k^i, h = 1, \dots, n_P^i \quad (14)$$

Among the  $n_P^i$  possible translations, we first select  $n_h$  ( $n_h < n_P^i$ ) translations which maximize the number of inliers. An inlier is defined as a point of  $P_{T_k}^i$  whose distance to a point of  $P_c^{i+1}$  is less than 5 pixels. However, due to the high variability that can generally be observed between two consecutive images, selecting a translation based only on the maximization of the number of inliers would not give optimal results. Hence, we propose to introduce additional information.

We first compute the region representing the tomato in the  $i^{\text{th}}$  image using a classical region growing algorithm, with seed calculated from the brightest pixels and



growth limited to the region inside  $\text{Ell}_f^i$ . We thus obtain the region  $\omega_t^i$  representing the non-occluded part of the tomato in the previous image (Fig. 10). This region is used as a reference for computing the optimal translation.

We then apply the region growing algorithm in the  $(i+1)^{\text{th}}$  image, considering every preselected translation (translation of seed and of the limiting area). We denote by  $\omega_k^{i+1}$  the binary image so obtained corresponding to a particular  $T_k$ .

Let us define  $v_k$ , the ratio of the two regions as:

$$v_k = \frac{|\omega_k^{i+1}|}{|\omega_t^i|} \quad (15)$$

where  $|A|$  represents the cardinality of set A. We retain the translation  $T_f$  (Fig. 11) which maximizes  $v_k$ , among the translations that satisfy two preliminary conditions:

$$v_k > 0.65 \quad (16)$$

$$\frac{1}{|\omega_k^{i+1}|} \sum_{(x,y) \in \omega_k^{i+1}} \text{Im}^{i+1}(x,y) > 0.65 \frac{1}{|\omega_t^i|} \sum_{(x,y) \in \omega_t^i} \text{Im}^i(x,y) \quad (17)$$

The first condition eliminates those cases where there is an inconsistency between the size of the region representing the tomatoes in the two consecutive images. The second condition ensures that the mean gray level measured in the  $(i+1)^{\text{th}}$  image on an area which is supposed to be a non-occluded part of the tomato is consistent with the one found in the previous image.



**Fig. 10**  $\omega_t^i$ : region representing tomato in  $i^{\text{th}}$  image

Due to the varying configurations of occlusion, it might be possible that none of the  $n_h$  preselected translations, which maximize the number of inliers, represent the actual translation of the tomato. This scenario is generally detected by the above two conditions. In that case, all the  $n_p^i$  initial translations are tested. We denote by  $C_t^{i+1}$  the updated location of the tomato center in the  $(i+1)^{\text{th}}$  image:

$$C_t^{i+1} = C_f^i + T_f \quad (18)$$

### 4.3 Estimation of the tomato boundary

In order to reduce the region to be analyzed, a smaller image  $\text{ImS}^{i+1}$  is extracted from  $\text{Im}^{i+1}$  with its center as  $C_t^{i+1}$ . A contrast stretching transformation is then applied to  $\text{ImS}^{i+1}$ .

The segmentation procedure is based on our active contour model with elliptic shape constraint (Section 3.3). It leads to a better robustness with respect to occlusions and lack of contrast. However, given the complexity of the scene, a good initialization is required; otherwise, the active contour will not converge towards the searched contours but will get trapped in another local minimum of the energy functional. Thus, the tomato detection procedure presented in Section 4.2 is of major interest as compared to our previous work [7, 8], as it improves the accuracy of the initialization algorithm (Section 4.4.1) and consequently the global performances (Section 4.4.2).

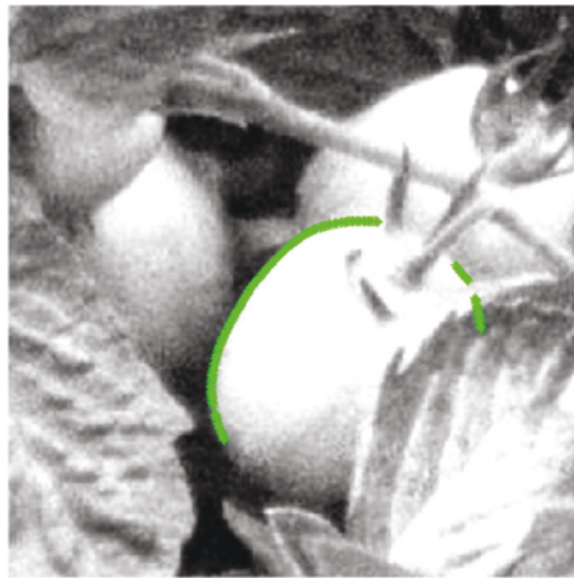
A brief overview of the initialization procedure is presented in Section 4.3.1. More details can be found in [7, 8]. Then, additional information about the active contour model is given in Section 4.3.2.

#### 4.3.1 Initialization of the active contour model

We use both gradient and region information to determine the initial position of the active contour model.

First, the method described in Section 4.2.2 (Eqs. 11 and 12) is applied in order to determine a set of candidate points, based on gradient magnitude and direction. Since the position of the tomato center is detected more accurately than in our previous work, it has been possible to restrict the size of the region of interest, which now lies inside two concentric circles with radius  $0.9b_f^i$  and  $1.1a_f^i$  instead of  $0.5r_f^i$  and  $1.5r_f^i$  (Fig. 12).

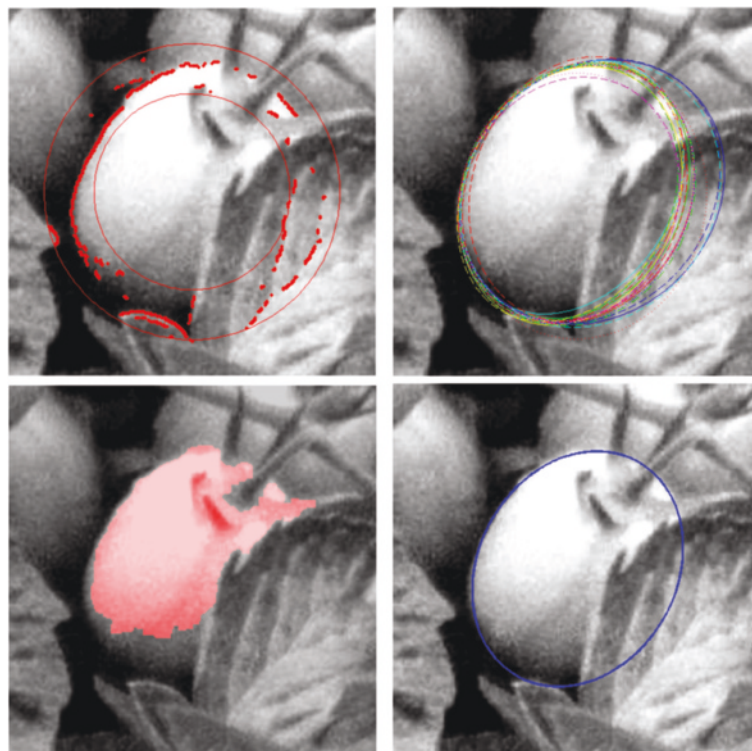
Due to the presence of outliers, even small, a least square estimate of an ellipse from all of the candidate points would not accurately represent the tomato boundary. Consequently, a RANSAC estimate based on an elliptic model is used to determine the parameters of several candidate ellipses. In this step, only those ellipses which satisfy the temporal constraint formulated in Section 3.2 are considered. Thus, a total of  $N_a = 20$  ellipses,  $\text{Ell}_u^{i+1}, u = 1, \dots, N_a$ , are retained: the ellipses



**Fig. 11** Translated points

with the largest number of inliers and whose parameters are compatible with the ones of the tomato in the previous image (Fig. 12). Note that both spatial regularization and temporal regularization have been used in this step, increasing the reliability of the segmentation procedure.

The third step of the initialization procedure aims at introducing region information, in order to select the best ellipse among the  $N_a$  previously retained. For that, we apply a region growing algorithm, whose seed is formed from the brightest pixels inside the intersection of all



**Fig. 12** Initialization of the elliptic active contour model. (*upper left*) Selection of candidate contour points; (*upper right*) selection of  $N_a$  ellipses, through RANSAC algorithm; (*lower left*) detection of the non-occluded parts of the tomato,  $\omega_t^{i+1}$ , through region growing algorithm; (*lower right*) initialization of the elliptic active contour model

the candidate ellipses. The result is a binary image  $\omega_t^{i+1}$  (Fig. 12), where the pixels set to 1 correspond to non-occluded parts of the tomato. Then, for every candidate ellipse  $\text{Ell}_u^{i+1}$ ,  $u = 1, \dots, N_a$ , we compare the region  $\omega_t^{i+1}$  with the region inside the considered ellipse, denoted by  $\omega_u^{i+1}$ , by calculating the following index:

$$\tau(u) = \frac{|\omega_u^{i+1} \cap (1 - \omega_t^{i+1})| + |\omega_t^{i+1} \cap (1 - \omega_u^{i+1})|}{|\omega_t^{i+1} \cap \omega_u^{i+1}|} \quad (19)$$

where  $|A|$  represents the cardinality of a set  $A$ . The ratio  $\tau(u)$  measures the consistency between the segmentation obtained through the contour analysis ( $\omega_u^{i+1}$ ) and the region analysis ( $\omega_t^{i+1}$ ). It reaches a minimum (zero) when  $\omega_u^{i+1}$  and  $\omega_t^{i+1}$  match perfectly.

Let us denote by  $a_u^{i+1}$  and  $b_u^{i+1}$  the semi axis lengths of the candidate ellipse  $\text{Ell}_u^{i+1}$ . The final elliptic initialization  $\text{Ell}_v^{i+1}$  (Fig. 12) is selected among the ellipses for which a good match is achieved (Eq. 20), considering another regularization condition which imposes that the size and shape of the ellipse in  $\text{ImS}^{i+1}$  are close to the ones in  $\text{ImS}^i$  (Eq. 21):

$$S_\tau = \left\{ u \in [1, N_a] \mid \tau(u) \leq 1.1 \min_{w \in [1, N_a]} \tau(w) \right\} \quad (20)$$

$$v = \arg \min_{u \in S_\tau} (a_u^{i+1} - a^i)^2 + (b_u^{i+1} - b^i)^2 \quad (21)$$

Combining  $\omega_v^{i+1}$  and  $\omega_t^{i+1}$  enables us also to determine the region of potential occlusions.

#### 4.3.2 Applying the elliptic active contour model

The elliptic active contour model (Section 3.3) is then applied from the initialization  $\text{Ell}_v^{i+1}$ . During the first  $n_{\text{start}}$  iterations, the parameter  $\psi$  is set to zero, so that  $z$  moves towards the most prominent contours. Then, the elliptic shape constraint is introduced for  $n_{\text{ellipse}}$  iterations ( $\psi \neq 0$ ) in order to guarantee robustness with respect to occlusion. Finally, the shape constraint is relaxed ( $\psi = 0$ ) for a few  $n_{\text{end}}$  iterations, which allows reaching the boundary more accurately, as a tomato is not a perfect ellipse (Fig. 13).

During the  $n_{\text{ellipse}}$  iterations, the reference ellipse is regularly and automatically updated from the current curve  $z$ . Again, 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 a procedure, similar to the one described in Section 4.2.1, is applied in order to select a subset of points that are very likely to lie on the boundary of the tomato. From these points, the parameters of the reference ellipse are optimized in a root

mean square error sense and so automatically updated every 10 iterations. Note that the length of the major and minor axes are estimated only once, at the beginning of the process, as the initial values are supposed to be very close to the actual values (temporal regularization), contrary to the other parameters of the ellipse, which are much more unstable due to the global movement of the tomato.

It is also worth noting that the image forces are not considered in the regions of occlusion, in every step of this process.

#### 4.3.3 Final elliptic approximation

Finally, four elliptic estimates (Fig. 13) of the tomato boundary are determined: points that are likely to be on the actual boundary are extracted based on several selection criteria [7, 8]; then, the RANSAC algorithm or a least square estimation [14] are applied to get the four elliptic approximations from the sets of selected points. In general, the four ellipses are almost the same in the case of little occlusion while they may differ more significantly in the case of higher occlusion. The operator has only to select the best estimate, if he considers it correct, or manually define a better elliptic approximation, otherwise. The latter case is however very rare and arises when the tomato is highly occluded (see the experimental results presented in Section 4.4).

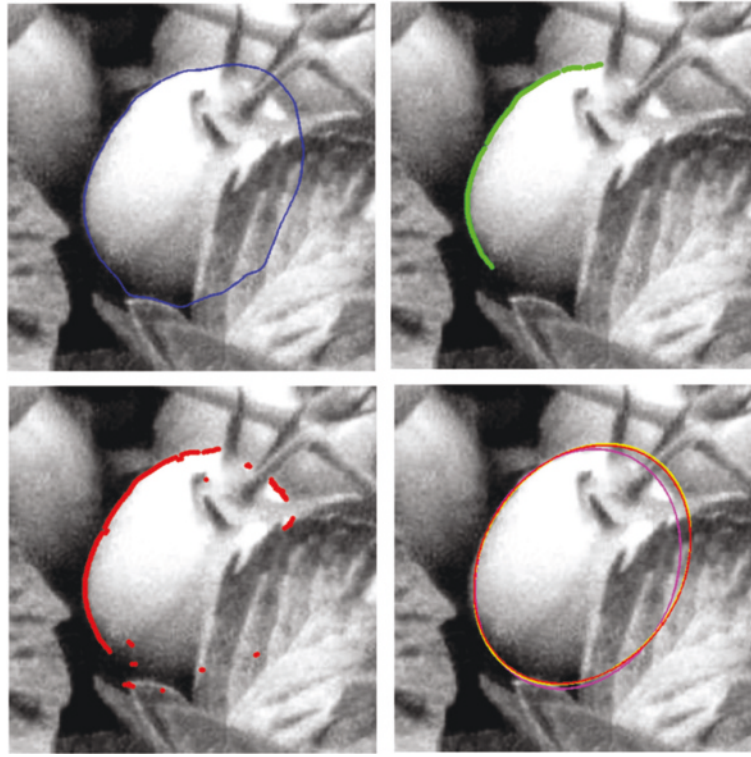
#### 4.4 Evaluation of the segmentation procedure

We first compare the proposed descriptor-based method to update the position of the tomato (Section 4.4.1) with our previous work. We then evaluate the segmentation procedure by comparing the obtained segmentations with manual segmentations (Section 4.4.2).

The proposed segmentation procedure was evaluated on the images acquired during three agriculture seasons (April–August, 2011–2013). Although a variety of the tomatoes was the same, a difference in vegetation was observed due to external climatic conditions. We identified 21 tomatoes for our study covering different sites and different seasons, thus ensuring a good representation of the variability. Analyzing only one pair of images a day for each tomato, we created therefore 21 pairs of image sequences.

Not all flowers develop into tomatoes at the same time. Besides, some tomatoes may be totally hidden by other tomatoes or leaves. As a result, the total number of days a particular tomato can be observed is not identical for all the 21 tomatoes. Flowers that mature to tomato early and are not hidden by leaves/tomatoes can be observed for a maximum number of days, thus creating a maximum number of images in the corresponding tomato sequence.

It is difficult to evaluate the segmentation procedure on the entire image dataset, given the variable degrees



**Fig. 13** Tomato segmentation. (upper left) Result provided by the elliptic active contour model; (upper right)(lower left) two sets of contours points extracted from (upper left); (lower right) four elliptic approximations

of occlusion. Therefore, the influence of the amount of occlusion on the final radius estimate was studied, and the image dataset was then divided into three categories [7, 8]:

$$\text{category} = \begin{cases} 1, & \text{if the amount of occlusion is less than 30 \%} \\ 2, & \text{if the amount of occlusion is between 30 and 50 \%} \\ 3, & \text{if the amount of occlusion is more than 50 \%} \end{cases} \quad (22)$$

Our evaluation is based on the comparison of the automatic segmentation results with elliptic approximation of manual segmentations. We denote the manual segmentation of the  $i^{\text{th}}$  image of a given sequence by  $\text{Ell}_s^i = [xc_s^i, yc_s^i, a_s^i, b_s^i, \phi_s^i]$  where  $C_s^i = [xc_s^i, yc_s^i]$  represents the center of the ellipse whose semi major and minor axes are  $a_s^i$  and  $b_s^i$ , respectively, and which has a rotation angle of  $\phi_s^i$ . Results are presented for category 1 and category 2 separately. Note that only images acquired from the left camera are presented in this section. Indeed, the images acquired using the left and the right cameras exhibit similar characteristics, in overall, even if different percentages of occlusion can be observed for some pairs.

#### 4.4.1 Tomato localization

In this section, we compare the proposed descriptor-based approach to update the position of the tomato with

the pattern matching approach presented in our previous work [7, 8]. For this, we measure the distance between the estimated center of the tomato obtained with the two approaches with the actual center of the tomato given by the manual segmentation.

Let us define for any image  $i$  two distance measures  $D_{\text{pm}}^i$  and  $D_{\text{desc}}^i$ :

$$D_{\text{pm}}^i = d(C_m^i, C_s^i) \quad (23)$$

$$D_{\text{desc}}^i = d(C_t^i, C_s^i) \quad (24)$$

$D_{\text{pm}}^i$  and  $D_{\text{desc}}^i$  represent the error on the estimation of the tomato center for the pattern matching method ( $C_m^i$ ) and the descriptor-based method ( $C_t^i$ ), respectively. Tables 1 and 2 show the percentage of images in a given sequence for which the distance measure  $D_{\text{pm}}^i$  or  $D_{\text{desc}}^i$  is less than 10 pixels,  $\frac{r_s^i}{4}$  and  $\frac{r_s^i}{2}$ , where  $r_s^i = \frac{a_s^i + b_s^i}{2}$ .

For the images of category 1, the position of the tomato was precisely detected in 97 % of the images using the descriptor-based approach ( $D_{\text{desc}}^i < 10$  pixels) as compared to 65 % in case of pattern matching. This demonstrates a significant improvement regarding the accuracy of the tomato localization. Moreover, almost

**Table 1** Percentages of images in category 1 where the distance measure  $D_{pm}^i$  or  $D_{desc}^i$  is less than a given threshold (10,  $\frac{r_s^i}{2}$  and  $\frac{r_s^i}{4}$  pixels). Also shown is the total number of images ( $N_1$ ) for each sequence in category 1

		Percentage of images with						
		$N_1$	$D_{desc} < 10$	$D_{pm} < 10$	$D_{desc} < \frac{r_s}{4}$	$D_{pm} < \frac{r_s}{4}$	$D_{desc} < \frac{r_s}{2}$	$D_{pm} < \frac{r_s}{2}$
Sequence	1	26	100.00	57.69	100.00	100.00	100.00	100.00
Sequence	2	4	100.00	100.00	100.00	100.00	100.00	100.00
Sequence	3	21	100.00	19.05	100.00	47.62	100.00	100.00
Sequence	4	14	92.86	78.57	100.00	100.00	100.00	100.00
Sequence	5	5	100.00	100.00	100.00	100.00	100.00	100.00
Sequence	6	0		–		–		–
Sequence	7	25	100.00	24.00	100.00	76.00	100.00	100.00
Sequence	8	20	100.00	60.00	100.00	90.00	100.00	100.00
Sequence	9	1	100.00	0	100.00	100.00	100.00	100.00
Sequence	10	5	100.00	20.00	100.00	100.00	100.00	100.00
Sequence	11	4	50.00	25.00	100.00	75.00	100.00	100.00
Sequence	12	19	100.00	84.21	100.00	94.74	100.00	94.74
Sequence	13	5	80.00	20.00	80.00	20.00	100.00	20.00
Sequence	14	4	100.00	100.00	100.00	100.00	100.00	100.00
Sequence	15	0		–		–		–
Sequence	16	21	100.00	95.24	100.00	95.24	100.00	100.00
Sequence	17	20	95.00	100.00	95.00	95.00	95.00	100.00
Sequence	18	23	86.96	91.3	86.96	91.3	86.96	91.3
Sequence	19	0		–		–		–
Sequence	20	5	100.00	100.00	100.00	100.00	100.00	100.00
Sequence	21	25	100.00	100.00	100.00	100.00	100.00	100.00
		247	96.76	65.28	97.98	88.05	98.38	94.78

all tomatoes (98 %) are correctly detected ( $D_{desc}^i < \frac{r_s^i}{2}$ ) based on this method with a significant improvement (+3.6 %) as compared with the pattern matching approach.

The images of category 2 contain a significant amount of occlusion, with more than 30 % of the elliptical contour hidden. Therefore, the pattern matching approach fails to find the position of the tomato ( $D_{pm}^i > \frac{r_s^i}{2}$ ) in 9 % of cases. However, the descriptor-based approach correctly detects the position of the tomato for 96.5 % of the images with a significant improvement (+5.1 %). For instance, in sequence 13, the position of the tomato was correctly detected ( $D_{pm}^i < \frac{r_s^i}{2}$ ) in only 76 % of the images by the pattern matching approach, against 100 % with the new one. Moreover, as in the case of images of category 1, a significant improvement regarding the accuracy of the estimated position of the tomato is observed. Using the descriptor-based approach, the position is accurately detected for 85 % of the images ( $D_{desc}^i < 10$  pixels), compared to 59 % with pattern matching.

The pattern matching approach [7, 8] is based on the detection of the non-occluded region. In case of partial occlusion, the maximum of correlation may be rather far from the actual center of the tomato. Moreover, it cannot provide an accurate estimation when several tomatoes overlap. The descriptor-based approach overcomes these difficulties since it relies on both region and contour information. The movement of the tomato can be accurately estimated by matching feature vectors calculated on contour points. Overall, the descriptor-based approach is more robust to occlusion and provides a more accurate estimation of the tomato center. This benefits to the complete segmentation procedure, as the detection of the candidate contour points is then much more reliable, leading to a better initialization of the elliptic active contour model.

#### 4.4.2 Tomato segmentation

In order to evaluate the segmentation procedure, the obtained segmentation  $A$  (one of the four final estimates  $Ell_{f1}^i$ ,  $Ell_{f2}^i$ ,  $Ell_{f3}^i$ , or  $Ell_{f4}^i$ ) is compared with the manual



**Table 2** Percentages of images in category 2 where the two distance measures  $D_{pm}^i, D_{desc}^i$  is less than a given threshold ( $10, \frac{r_s^i}{2}$  and  $\frac{r_s^i}{4}$  pixels). Also shown is the total number of images ( $N_2$ ) of category 2 in each sequence

		Percentage of images with						
		$N_2$	$D_{desc} < 10$	$D_{pm} < 10$	$D_{desc} < \frac{r_s}{4}$	$D_{pm} < \frac{r_s}{4}$	$D_{desc} < \frac{r_s}{2}$	$D_{pm} < \frac{r_s}{2}$
Sequence	1	2	100.00	100.00	100.00	100.00	100.00	100.00
Sequence	2	3	66.67	66.67	100.00	66.67	100.00	100.00
Sequence	3	16	93.75	31.25	100.00	56.25	100.00	87.5
Sequence	4	13	92.31	61.54	100.00	100.00	100.00	100.00
Sequence	5	12	75	25	83.33	50.00	100.00	100.00
Sequence	6	6	83.33	33.33	83.33	50.00	100.00	100.00
Sequence	7	12	50.00	25.00	66.67	50.00	83.33	83.33
Sequence	8	14	100.00	50.00	100.00	78.57	100.00	100.00
Sequence	9	14	100.00	14.29	100.00	35.71	100.00	92.86
Sequence	10	3	66.67	0	66.67	0	66.67	66.67
Sequence	11	5	20.00	0	40.00	20.00	100.00	80.00
Sequence	12	0		–		–		–
Sequence	13	17	94.12	64.71	94.12	64.71	100.00	76.47
Sequence	14	16	87.5	100.00	87.5	68.75	93.75	100.00
Sequence	15	13	100.00	100.00	92.31	92.31	100.00	100.00
Sequence	16	0		–		–		–
Sequence	17	1	100.00	100.00	100.00	100.00	100.00	100.00
Sequence	18	1	100.00	100.00	100.00	100.00	100.00	100.00
Sequence	19	10	90.00	50.00	90.00	50.00	90.00	50.00
Sequence	20	10	90.00	100.00	90.00	90.00	100.00	100.00
Sequence	21	1	100.00	100.00	100.00	100.00	100.00	100.00
		169	84.70	59.04	89.15	66.99	96.51	91.41

segmentation  $\text{Ell}_s^i$  by computing the average  $D_{\text{mean}}^i$  and the maximum  $D_{\text{max}}^i$  distance between  $A$  and  $\text{Ell}_s^i$  for every tomato image  $i$ . These distance measures, expressed in pixels, are normalized with respect to the size of the tomato ( $r_s^i$ ) in order to better interpret the results. The normalized distance measures are defined as:

$$D_{\text{meanR}}^i = \frac{D_{\text{mean}}^i}{r_s^i} 100 \quad (25)$$

$$D_{\text{maxR}}^i = \frac{D_{\text{max}}^i}{r_s^i} 100 \quad (26)$$

Tables 3 and 4 present the mean and the standard deviation of  $D_{\text{meanR}}^i$  and  $D_{\text{maxR}}^i$  for the images of category 1 and category 2, respectively. These tables do not consider the images with incorrect estimation of the tomato position (i.e.,  $D_{\text{desc}}^i > \frac{r_s^i}{2}$ ). Moreover, two series of results are presented: the statistics computed for  $\text{Ell}_{f4}^i$ , which would correspond to a fully automatic process, and the statistics computed with the best estimate  $\text{Ell}_{\text{opt}}^i$  (i.e. minimizing  $D_{\text{meanR}}^i$ ), which, in practice, would be selected by an operator among the four possibilities.

For the images of category 1, very good results (Fig. 14) were obtained in most of the sequences with  $\mu_{D_{\text{meanR}}}$  less than 10 % for all the sequences. The mean error averaged on all images is less than 5 % (even for  $\text{Ell}_{f4}$ , without manual selection), which fits the requirements of the final user. Moreover, a lower  $\sigma_{D_{\text{meanR}}}$  demonstrates the robustness of our method. Even if the occlusion degree is less than 30 %, many images present a blurred contour or low contrast, due to shadowing effects or overlap with nearby tomatoes. However, a good segmentation is obtained in even these challenging cases (Fig. 14b, c). For the images acquired in the agriculture season 2013 (Sequence 12–21), the size of the tomatoes is smaller compared to the seasons 2011 and 2012. This accounts for slightly higher relative distance measures for these sequences (Fig. 14d).

For the images of category 2, which contain a significant amount of occlusion (30 to 50 %), good results are observed with  $\mu_{D_{\text{meanR}}}$  less than 10 % in case of  $\text{Ell}_{f4}$  for almost all sequences (except sequence 10). When considering the optimal ellipse  $\text{Ell}_{\text{opt}}$ ,  $\mu_{D_{\text{meanR}}}$  is less than 10 % for all the sequences and less than 5 % for 60 %

**Table 3** Distribution of  $D_{\text{mean}R}^i$  and  $D_{\text{max}R}^i$  computed assuming  $\text{Ell}_{f4}^i$  or  $\text{Ell}_{\text{opt}}^i$  as the final segmentation for the images of category 1. Also shown are the total number of images ( $N_1$ ) of category 1 for each sequence and the number of images  $N_c$  where the position of the tomato is correctly estimated ( $D_{\text{desc}}^i < \frac{r_s^i}{2}$ ). Only four images were discarded using this criterion

	$N_1$	$N_c$	$\text{Ell}_{f4}$				$\text{Ell}_{\text{opt}}$			
			$\mu_{D_{\text{mean}R}}$	$\sigma_{D_{\text{mean}R}}$	$\mu_{D_{\text{max}R}}$	$\sigma_{D_{\text{max}R}}$	$\mu_{D_{\text{mean}R}}$	$\sigma_{D_{\text{mean}R}}$	$\mu_{D_{\text{max}R}}$	$\sigma_{D_{\text{max}R}}$
Sequence 1	26	26	2.86	1.81	8.73	5.93	1.79	1.22	5.57	4.1
Sequence 2	4	4	1.85	0.35	5.24	1.59	1.4	0.69	3.9	1.15
Sequence 3	21	21	1.28	2.68	10.99	6.49	3.46	2.43	9.45	7.48
Sequence 4	14	14	3.99	1.4	10.36	4.01	2.4	1.11	6.76	3.21
Sequence 5	5	5	3.51	1.5	10.42	3.82	3.15	1.2	8.87	2.69
Sequence 6	0	0	–	–	–	–	–	–	–	–
Sequence 7	25	25	1.89	0.88	5.45	3.09	1.48	0.55	4.43	1.84
Sequence 8	20	20	3.53	2.38	9.58	7.11	3.21	2.06	9.13	6.02
Sequence 9	1	1	2.87	0	9.03	0	2.61	0	8.55	0
Sequence 10	5	5	1.76	0.95	5.28	3.3	1.46	0.76	4.31	2.71
Sequence 11	4	4	4.6	3.94	15.39	14.65	4.06	3.73	14	14.03
Sequence 12	19	19	4.66	1.62	12.61	4.55	4.09	1.36	11.06	3.67
Sequence 13	5	5	9.88	5.98	15.87	15.24	9.72	5.98	25.87	15.9
Sequence 14	4	4	7.63	0.29	20.18	3.57	6.24	1.74	15.57	2.79
Sequence 15	0	0	–	–	–	–	–	–	–	–
Sequence 16	21	21	3.63	1.21	8.83	2.46	3.46	1.14	8.51	2.55
Sequence 17	20	19	8.12	2.53	17.62	5	7.79	2.41	17.01	4.45
Sequence 18	23	20	6.23	1.75	17.91	6.58	5.38	1.65	15.79	6.61
Sequence 19	0	0	–	–	–	–	–	–	–	–
Sequence 20	5	5	6.44	5.42	19.07	18.98	6.12	5.34	17.47	16.49
Sequence 21	25	25	3.75	1.42	9.38	2.79	3.65	1.38	9.38	2.79
	247	243	4.29	2.77	11.35	7.39	3.97	1.93	9.95	7.06

of them. The results have been significantly improved compared to our previous work, thanks to the more accurate detection of the position of the tomato. Figure 15 shows some examples where a good segmentation is obtained even in the presence of noise and blurred contour and/or low resolution, in addition to severe occlusion.

The significant segmentation errors occur when the presence of the neighboring leaves and/or branches produces strong gradients near the desired contour which misleads the movement of the active contour. This effect is prominent in sequence 13, which results in higher distance measures for this sequence (Fig. 16a). Other typical cases of errors result from blurring due to tomato overlap combined with partial occlusion (Fig. 16b).

Finally, it is worth noting that it may be very difficult to determine the contour position at some places even manually, because of blurring, shadowing (Fig. 17a, b) or the presence of leaves near the “head” of the tomato at the beginning of the maturation (Fig. 17c). This

imprecision can also explain middling results in some cases.

Table 5 compares the statistics obtained for the two methods, the one presented in [7, 8] ( $\mu_{\text{mean}R}^o, \sigma_{D_{\text{mean}R}}^o$ ) and the new one proposed in this paper ( $\mu_{\text{mean}R}, \sigma_{D_{\text{mean}R}}$ ). For each method, the images considered in the table are only the ones where the position of the tomato is correctly detected ( $D_{\text{pm}}^i < \frac{r_s^i}{2}, D_{\text{desc}}^i < \frac{r_s^i}{2}$ , respectively), since the segmentation algorithm does never provide correct results otherwise. In this way, more images ( $N_c > N_c^o$ ) can be processed with the descriptor-based approach, which shows better robustness with respect to occlusion (+8 % for images of category 2). The descriptor-based approach provides a significant benefit when the image quality is poor or the tomato is highly occluded (e.g., sequences 3, 4, 10, 11, 14, 15, 17, 20, 21). Similar results are obtained otherwise. This detailed study demonstrates that the update of the tomato position is a crucial step which conditions the quality of the final segmentation.

**Table 4** Distribution of  $D_{\text{meanR}}^i$  and  $D_{\text{maxR}}^i$  computed assuming  $\text{Ell}_{f4}^i$  and  $\text{Ell}_{\text{opt}}^i$  as the final segmentation for the images of category 2. Also shown are the total number of images ( $N_2$ ) of category 2 for each sequence and the number of images  $N_c$  where the position of the tomato is correctly estimated ( $D_{\text{desc}}^i < \frac{r_i^i}{2}$ ). Only five images were discarded using this criterion

	$N_2$	$N_c$	$\text{Ell}_{f4}$				$\text{Ell}_{\text{opt}}$			
			$\mu_{D_{\text{meanR}}}$	$\sigma_{D_{\text{meanR}}}$	$\mu_{D_{\text{maxR}}}$	$\sigma_{D_{\text{maxR}}}$	$\mu_{D_{\text{meanR}}}$	$\sigma_{D_{\text{meanR}}}$	$\mu_{D_{\text{maxR}}}$	$\sigma_{D_{\text{maxR}}}$
Sequence 1	2	2	4.61	2.02	14.45	7.13	2.14	0.57	4.93	1.85
Sequence 2	3	3	2.32	0.91	5.4	2.42	1.84	0.81	5.24	2.51
Sequence 3	16	16	6.54	4.9	17.85	13.42	5.22	4	15.42	13.01
Sequence 4	13	13	4.77	2.18	13.66	7.19	4.39	2.17	12.55	7.27
Sequence 5	12	12	4.59	3.11	14.31	9.23	3.37	2.13	11.26	7.75
Sequence 6	6	6	2.64	1.7	7.75	4.76	2.49	1.71	7.1	4.7
Sequence 7	12	10	6.49	5.04	18.12	15.27	4.27	4.55	12.71	13.79
Sequence 8	14	14	5.4	2.1	15.13	7.18	4.35	1.95	12.96	7.45
Sequence 9	14	14	6.68	3.99	17.18	9.63	5.06	4.07	13.32	10.61
Sequence 10	3	2	12.18	5.64	31.36	9.57	1.82	5.85	31.37	8.73
Sequence 11	5	5	5.41	2.42	19.39	10.84	4.86	2.66	17.6	12.5
Sequence 12	0	0	–	–	–	–	–	–	–	–
Sequence 13	17	17	9.19	4.54	23.59	11.65	7.76	3.51	22.9	12.77
Sequence 14	16	15	8.28	3.31	21.56	10.12	7.23	3.17	21.15	11.31
Sequence 15	13	13	3.57	1.84	9.7	4.55	3.04	1.4	8.42	3.74
Sequence 16	0	0	–	–	–	–	–	–	–	–
Sequence 17	1	1	7.71	0	22.23	0	7.36	0	22.23	0
Sequence 18	1	1	6.78	0	17.4	0	6.78	0	17.4	0
Sequence 19	10	9	5.76	4.13	14.2	10.63	4.85	3.28	11.8	7.31
Sequence 20	10	10	5.8	2.84	14.48	6.16	4.64	1.54	11.01	3.21
Sequence 21	1	1	2.95	0	9.36	0	2.79	0	8.88	0
	169	164	6.06	3.84	16.41	10.37	4.43	2.28	14.27	10.46

## 5 Size estimation

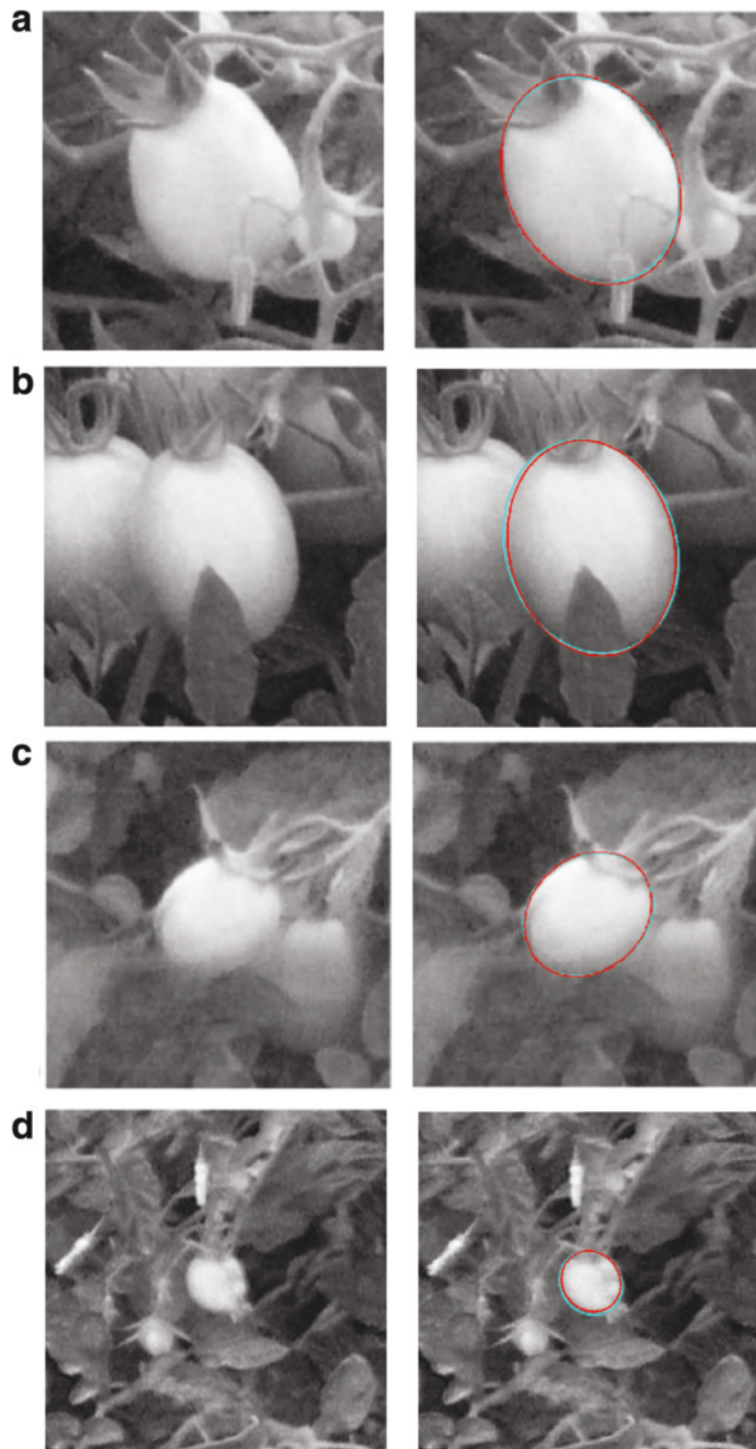
In this section, we wish to estimate the size of the tomato approximated by a sphere in the 3D space. In order to calculate the parameters of the sphere, it is assumed that the camera projection matrices,  $\mathbf{P}$  and  $\mathbf{Q}$  for the left and right cameras respectively, as well as the parameters of the apparent contours (ellipse) in the two images, have been calculated.

The camera parameters are determined once, at the beginning of the season, by observing a calibration pattern at different positions and orientations in the scene, as described in [9]. Note that the acquisition system is firmly fixed to the ground, so that the calculated projection matrices  $\mathbf{P}$  and  $\mathbf{Q}$  are valid throughout the agricultural season. The parameters of the apparent contours of the tomatoes in the left and right images are provided by the segmentation procedure, after validation by an operator (Section 4.3.3).

In the following, we use bold letters to denote vectors ( $\mathbf{X}$ ) and italics to denote scalars ( $X$ ). Quantities in

the 3D space are denoted by upper case letters ( $\mathbf{X}$ ,  $X$ ) while image quantities are denoted by lower case letters ( $\mathbf{x}$ ,  $x$ ). Finally, points lying on the ellipses obtained from the segmentation algorithm are now denoted by  $\mathbf{x}_{o,l}$  and  $\mathbf{x}_{o,r}$ ,  $o = 1, \dots, N$ , for the left and right images respectively.

From the segmentation in the two images, we first recover the ellipse centers,  $\mathbf{x}_l$  in the left image and  $\mathbf{x}_r$  in the right image. It is assumed that these are the image points of the sphere center  $\mathbf{X}_{ct}$ , which is then determined based on a triangulation procedure [9]. Using the property presented in Section 3.1, a set of 3D space points lying on the contour generator is determined from the points on the elliptic contour, for each image (Section 5.1). From the two sets of 3D space points, corresponding to the left and right images, two values for the radius of the sphere are computed based on a least square minimization approach (Section 5.2). Finally, a joint functional is minimized in order to obtain an estimate of the sphere radius (Section 5.3).

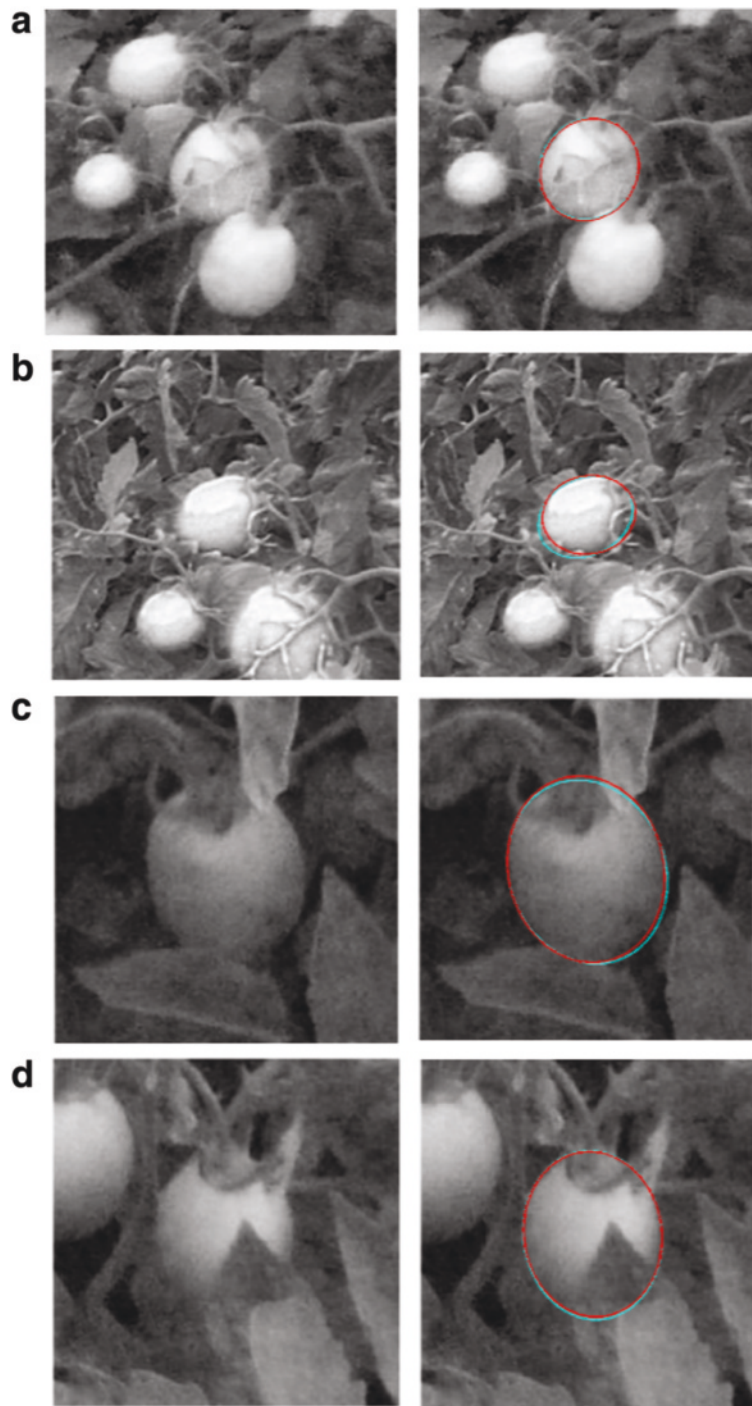


**Fig. 14** Original image (left) and final segmentation (right) shown in red obtained on images of category 1. The contour in cyan represents the manual segmentation. The distance measures ( $D_{\text{meanR}}$ ,  $D_{\text{maxR}}$ ) are (top to bottom) (1.06 %, 2.98 %), (2.12 %, 4.79 %), (1.12 %, 3.25 %), (5.08 %, 13.38 %). Note that even in the presence of occlusion (a), smoothed contour (b, c), a good segmentation is obtained. Also notice the small size of the tomato which results in higher relative distance measure (d)

### 5.1 3D space points lying on the contour generator

Let us first consider the left image and the image points  $\mathbf{x}_{o,l}$ ,  $o = 1, \dots, N$ . The 3D space points lying on the

contour generator are computed as the intersection of the rays back projected from the image points and the plane  $\Pi_l$  which contains the sphere center  $\mathbf{X}_{ct}$  and is orthogonal



**Fig. 15** Original image (left) and final segmentation (right) shown in red obtained on four (a-d) images of category 2. The contour in cyan represents the manual segmentation. The distance measures ( $D_{\text{meanR}}$ ,  $D_{\text{maxR}}$ ) are (top to bottom) (1.25 %, 2.74 %), (5.50 %, 17.71 %), (2.08 %, 5.89 %), (1.20 %, 3.74 %)

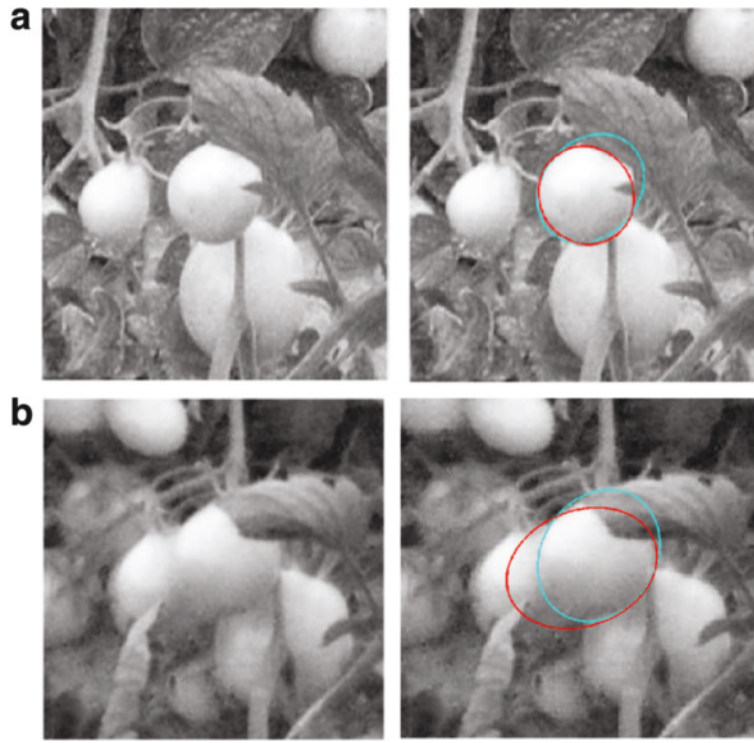
to the line joining the camera center  $\mathbf{C}_l$  (deduced from the projection matrix  $\mathbf{P}$ ) and  $\mathbf{X}_{ct}$ . Its equation is given by:

$$AX + BY + CZ + D = 0, \quad (27)$$

where

$$N = [A, B, C]^T = \frac{\mathbf{X}_{ct} - \mathbf{C}_l}{\|\mathbf{X}_{ct} - \mathbf{C}_l\|}, D = -AX_{ct} - BY_{ct} - CZ_{ct}. \quad (28)$$





**Fig. 16** Original image (left) and final segmentation (right) shown in red obtained on two (a-b) images of sequence 13. The contour in cyan represents the manual segmentation. The distance measures ( $D_{\text{meanR}}$ ,  $D_{\text{maxR}}$ ) are (top to bottom) (10.39 %, 37.76 %), (19.87 %, 50.23 %)

The 3D space point  $\mathbf{X}_{o,l} = [X_{o,l}, Y_{o,l}, Z_{o,l}]^T$  is projected to its corresponding image point  $\mathbf{x}_{o,l} = [x_{o,l}, y_{o,l}]^T$  via the projection matrix  $\mathbf{P}$  as:

$$\beta [x_{o,l} \ y_{o,l} \ 1]^T = \mathbf{P} [X_{o,l} \ Y_{o,l} \ Z_{o,l} \ 1]^T \quad (29)$$

where  $\beta$  is an unknown non-zero scalar factor. By expanding the above equation and eliminating the unknown  $\beta$ , two equations are obtained in the unknowns  $X_{o,l}, Y_{o,l}, Z_{o,l}$ . Moreover, it is assumed that  $\mathbf{X}_{o,l}$  lies on plane  $\Pi_l$  (Eq. 27). This gives a set of three equations which may be written in the form:

$$\mathbf{A}_p \mathbf{X}_{o,l} = \mathbf{B}_p \quad (30)$$

where,

$$\mathbf{A}_p = \begin{bmatrix} x_{o,l}P_{31} - P_{11} & x_{o,l}P_{32} - P_{12} & x_{o,l}P_{33} - P_{13} \\ y_{o,l}P_{31} - P_{21} & y_{o,l}P_{32} - P_{22} & y_{o,l}P_{33} - P_{23} \\ A & B & C \end{bmatrix} \quad (31)$$

$$\mathbf{B}_p = \begin{bmatrix} P_{14} - P_{34}x_{o,l} \\ P_{24} - P_{34}y_{o,l} \\ -D \end{bmatrix} \quad (32)$$

Assuming  $\mathbf{A}_p$  to be non-singular, the solution of this equation is given by:

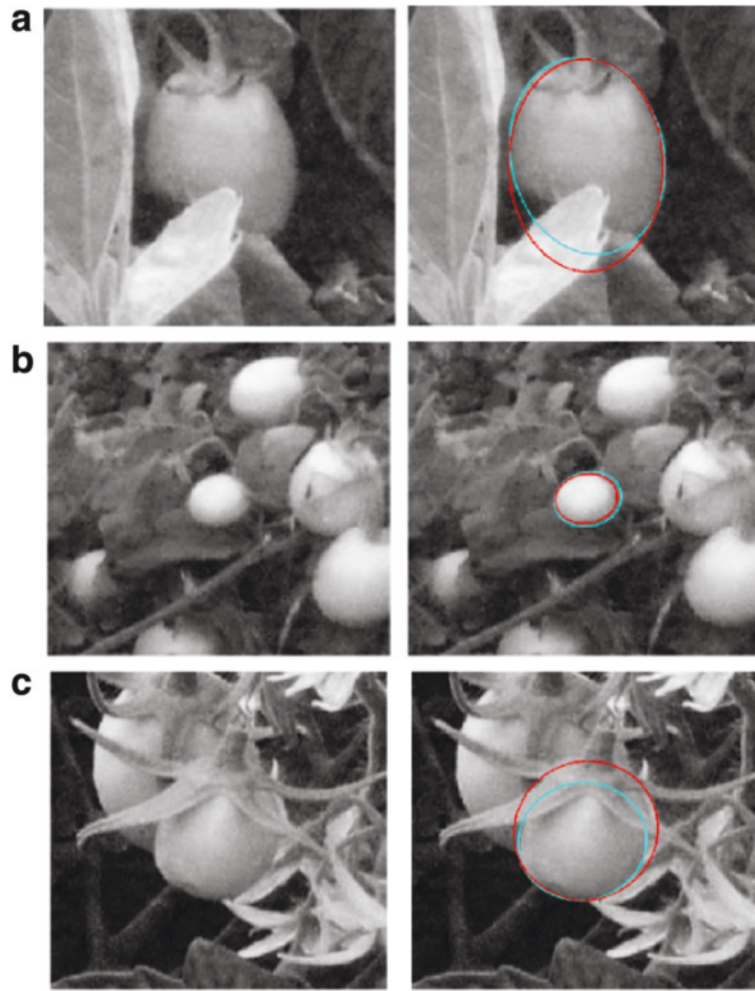
$$\mathbf{X}_{o,l} = \mathbf{A}_p^{-1} \mathbf{B}_p \quad (33)$$

The same process is performed for the right image. Consequently, two sets of 3D space points  $\mathbf{X}_{o,l}$  and  $\mathbf{X}_{o,r}$  are computed from their respective image points  $\mathbf{x}_{o,l}$  and  $\mathbf{x}_{o,r}$  in the left and right images.

## 5.2 Least square estimation of the circle

This estimation is performed on the left and right images independently. Let us consider the left image. Due to measurement errors, the 3D space points  $\mathbf{X}_{o,l}$ ,  $o = 1, \dots, N$  might not exactly lie on a perfect circle in the plane  $\Pi_l$ . Hence, we search for a least square estimate of the circle.

In order to simplify the calculation, every 3D space point  $\mathbf{X}_{o,l}$  is transformed to a new coordinate system  $(X', Y', Z')$  linked to the plane  $\Pi_l$  with  $X', Y'$  axes lying on the plane and  $Z'$ -axis orthogonal to the plane. The least square estimation of the circle enables us to get a first estimate of the sphere radius,  $R_l$ . We get the second estimate  $R_r$  in the same way.



**Fig. 17** Original image (left) and final segmentation (right) shown in red for three images (a-c). The contour in cyan represents the manual segmentation. The distance measures ( $D_{\text{mean}R}$ ,  $D_{\text{max}R}$ ) are (top to bottom) (6.54 %, 21.97 %), (7.92 %, 14.46 %), (15.95 %, 37.54 %)

### 5.3 Joint optimization

Let us denote by  $[X'_{ct,l}, Y'_{ct,l}]$  ( $[X'_{ct,r}, Y'_{ct,r}]$ ) the center of the sphere in the left image (right image) in the new coordinate system linked to  $\Pi_l$  ( $\Pi_r$ ). From an initial radius estimate defined as the mean of  $R_l$  and  $R_r$  (i.e.  $R = \frac{R_l + R_r}{2}$ ), the following function:

$$F(R) = \sum_{o=1}^N \left[ \left( X'_{o,l} - X'_{ct,l} \right)^2 + \left( Y'_{o,l} - Y'_{ct,l} \right)^2 - R^2 \right]^2 + \left[ \left( X'_{o,r} - X'_{ct,r} \right)^2 + \left( Y'_{o,r} - Y'_{ct,r} \right)^2 - R^2 \right]^2 \quad (34)$$

is minimized using the Gauss-Newton method, in order to determine the final estimate of the sphere radius, denoted by  $R_{\text{est}}$ .

### 5.4 Evaluation: Size estimation

Since a tomato is not a perfect sphere, two reference distances  $D_1, D_2$  were measured manually, which approximate the size of the tomato (Fig. 18). These reference distances were compared with the estimated radius  $R_{\text{est}}$  and relative error percentages  $PE_{D_1}, PE_{D_2}$  were computed with respect to  $D_1$  and  $D_2$ :

$$PE_{D_1} = 100 \left| \frac{D_1 - R_{\text{est}}}{D_1} \right| \quad (35)$$

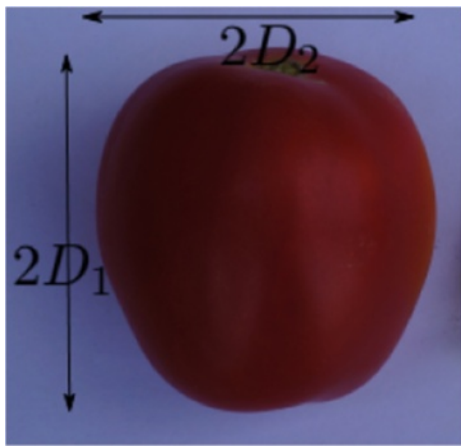
$$PE_{D_2} = 100 \left| \frac{D_2 - R_{\text{est}}}{D_2} \right| \quad (36)$$

In order to evaluate the proposed method, the radius of some tomatoes ( $T_a=1, \dots, 10$ ) were measured from images acquired in the laboratory under ideal conditions (correct illumination, no occlusion). Images of these tomatoes were acquired at different positions (Pos A and B) and different heights ( $H = 10$  and  $30$  cm) from the ground using

**Table 5** Comparing the mean and standard deviation of  $D_{\text{meanR}}$  for the two methods in the images of categories 1 and 2 (Ell<sub>opt</sub>) where the position of the tomato is correctly estimated

Sequence	Category 1								Category 2							
	$N_1$	$N_C^0$	$\mu_{D_{\text{meanR}}}^0$	$\sigma_{D_{\text{meanR}}}^0$	$N_C$	$\mu_{D_{\text{meanR}}}$	$\sigma_{D_{\text{meanR}}}$	Difference	$N_2$	$N_C^0$	$\mu_{D_{\text{meanR}}}^0$	$\sigma_{D_{\text{meanR}}}^0$	$N_C$	$\mu_{D_{\text{meanR}}}$	$\sigma_{D_{\text{meanR}}}$	Difference
1	26	26	1.34	0.7	26	1.79	1.22	-0.45	2	2	1.9	0.47	2	2.14	0.57	-0.24
2	4	4	1.57	0.46	4	1.4	0.69	0.17	3	3	1.57	0.57	3	1.84	0.81	-0.27
3	21	21	2.87	2.1	21	3.46	2.43	-0.59	16	14	7.36	8.03	16	5.22	4	2.14
4	14	14	2.2	1.84	14	2.4	1.11	-0.2	13	12	6.21	5.55	13	4.39	2.17	1.82
5	5	5	4.54	1.28	5	3.15	1.2	1.39	12	12	3.48	1.63	12	3.37	2.13	0.11
6	0	0	-	-	0	-	-		6	6	4.63	3.75	6	2.49	1.71	2.14
7	25	25	1.7	0.5	25	1.48	0.55	0.22	12	10	4.5	4.52	10	4.27	4.55	0.23
8	20	20	5.4	5.01	20	3.21	2.06	2.19	14	14	4.73	2.38	14	4.35	1.95	0.38
9	1	1	5.24	0	1	2.61	0	2.63	14	13	5.07	4.08	14	5.06	4.07	0.01
10	5	5	1.75	0.4	5	1.46	0.76	0.29	3	2	5.15	2.98	2	1.82	5.85	3.33
11	4	4	10.18	6.21	4	4.06	3.73	6.12	5	4	8.95	0	5	4.86	2.66	4.09
12	19	18	4.35	1.34	19	4.09	1.36	0.26	0	0	-	-	0	-	-	
13	5	1	8.56	0	5	9.72	5.98	-1.16	17	13	9.26	4.83	17	7.76	3.51	1.5
14	4	4	9.18	2.96	4	6.24	1.74	2.94	16	16	11.44	7.62	15	7.23	3.17	4.21
15	0	0	-	-	0	-	-		13	13	5.93	8.07	13	3.04	1.4	2.89
16	21	21	4.46	1.21	21	3.46	1.14	1	0	0	-	-	0	-	-	
17	20	20	11.56	2.55	19	7.79	2.41	3.77	1	1	10.63	0	1	7.36	0	3.27
18	23	21	7.77	2.24	20	5.38	1.65	2.39	1	1	7.41	0	1	6.78	0	0.63
19	0	0	-	-	0	-	-		10	5	6.31	1.86	9	4.85	3.28	1.46
20	5	5	6.88	2.57	5	6.12	5.34	0.76	10	10	7.51	3.13	10	4.64	1.54	2.87
21	25	25	7.12	3.22	25	3.65	1.38	3.47	1	1	5.67	0	1	2.79	0	2.88
	247	240	5.37	1.92	243	3.97	1.93	1.40	169	152	6.20	3.13	164	4.43	2.28	1.76

the same acquisition system as in the open field. The relative position between the camera and tomatoes was also identical, providing images of similar resolution. At each position, the tomatoes were observed at three different orientations (Orn = 1, 2, 3), not necessarily identical for all positions. The radius of these tomatoes were computed

**Fig. 18** Reference distances  $D_1, D_2$ 

as described above using a manual segmentation as we wish to focus on the evaluation of the second part of the system (estimation of the tomato size). Table 6 shows the  $PE_{D_1}$  for different positions and different orientations. This error is always less than 10 % and most (91 %) of the values are less than 5 %, which demonstrates the robustness of the radius estimation. Moreover, the estimated radius  $R_{\text{est}}$  is closer to the reference distance  $D_1$  than  $D_2$ . This is because a sphere with a radius  $D_1$  will cover the entire tomato and hence it is logical that  $R_{\text{est}}$  is closer to  $D_1$ .

Using the estimated radius  $R_{\text{est}}$ , an estimate of the volume,  $V_{R_{\text{est}}}$ , is computed using the spherical hypothesis ( $V_{R_{\text{est}}} = \frac{4}{3}\pi R_{\text{est}}^3$ ). However, since a tomato is not a perfect sphere, we determined a correction factor  $\alpha_{cc}$  that can be applied on the radius in order to get a measure closer to the actual volume.

Let us denote by  $V_{D_1}^{\text{correc}}$  the volume estimated with the corrected radius  $R = \alpha_{cc}D_1$  (i.e.,  $V_{D_1}^{\text{correc}} = \frac{4}{3}\pi(\alpha_{cc}D_1)^3$ ). The value of  $\alpha_{cc}$  has been determined experimentally so as to minimize the relative difference  $\frac{|V_{\text{actual}} - V_{D_1}^{\text{correc}}|}{V_{\text{actual}}}$  for four ( $T_a = 1, 2, 3, 4$ ) tomatoes studied in all positions,

**Table 6** Percentage error  $PE_{D_1}$  for the tomatoes  $T_a$ 

Pos	$H$	Orn	$T_a$									
			1	2	3	4	5	6	7	8	9	10
A	10	1	2.11	2.68	4.79	1.79	3.27	1.42	4.70	4.03	4.20	6.63
		2	2.47	7.32	3.81	1.10	0.15	1.09	3.80	3.87	5.07	8.46
		3	4.51	5.65	4.66	0.33	1.98	0.94	2.66	5.02	5.01	5.65
	30	1	1.51	4.63	0.34	2.70	1.83	4.61	4.20	1.61	2.10	1.91
		2	0.58	3.97	5.32	0.95	0.86	0.04	2.19	1.14	0.25	0.30
		3	1.00	4.73	5.13	4.53	2.52	3.94	2.49	0.39	1.92	0.37
B	10	1	4.61	1.59	0.64	0.97	4.23	2.23	0.47	2.21	0.17	4.76
		2	4.89	0.87	0.03	0.42	2.35	1.74	0.03	1.29	2.62	0.88
		3	4.05	0.55	1.71	1.08	0.45	1.72	0.22	3.29	2.60	3.03
	30	1	0.47	1.75	1.69	0.92	0.30	1.11	3.47	1.60	2.38	3.54
		2	3.10	1.75	1.37	0.66	0.79	1.27	0.38	0.00	2.09	1.93
		3	1.62	1.02	0.65	2.43	0.41	1.17	6.07	4.50	0.39	0.74

where  $V_{\text{actual}}$  is the actual volume. We found  $\alpha_{cc} = 0.95$ . The relative error percentage  $PE_{V_{\text{actual}}}^{\text{correc}}$  between  $V_{R_{\text{est}}}^{\text{correc}}$  and the actual volume  $V_{\text{actual}}$  is then studied for the other tomatoes at different positions and different orientations. The error percentage is less than 15 % in 87 % of the cases. From this experiment, it seems that it may be possible to correct the measurements made with the spherical hypothesis in order to take into account the specific shape of the tomato that is cultivated in the field. This short study validates the volume estimation using the spherical hypothesis. However, the proposed correction model is very basic and was parametrized on a small set of tomatoes. Further studies need to be conducted on a larger dataset to increase the robustness of the volume estimation.

## 6 Result: entire system

The method proposed in Section 5 was used to measure the size of 10 tomatoes cultivated in the open field for the agriculture season 2013 (Sequences 12–21). It is assumed that  $\text{Ell}_{\text{opt}}$  is the final segmentation selected by the operator among the four possibilities, in both images (no manual correction). Table 7 presents the estimated

radius  $R_{\text{est}}^f$  along with the error percentage  $PE_{D_1}, PE_{D_2}$  with respect to the two reference distances  $D_1, D_2$  (Eqs. 35 and 36).

For 8 tomatoes over 10, the estimated radius is actually in the interval  $[D_2, D_1]$  and generally close to  $D_1$  as compared to  $D_2$  as discussed earlier. Moreover, for most of the sequences except 19, the error percentage  $PE_{D_1}$  is less than 10 %, which demonstrates the robustness of our method and fits the requirement of the project. For sequence 19, the position of the tomato is not correctly detected due the presence of identical neighboring tomatoes (Fig. 19). As a result, an incorrect radius is estimated for this tomato.

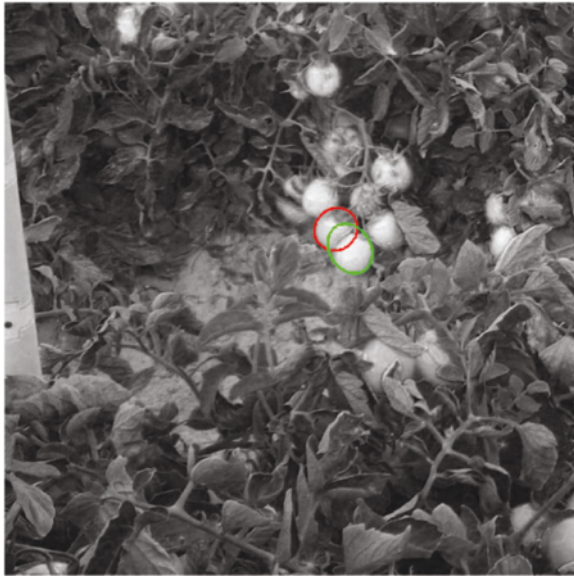
Note that the accuracy of the final radius estimates depends on the reliable estimation of several parameters at the different steps of the method (segmentation, camera parameters, etc.). An imprecision in one of these parameters would result in an inaccurate radius estimate. The influence of the imprecision in the segmentation on the estimated radius was also studied theoretically using Eq. 1. For a 1-pixel error in the length of the major axis (minor axis) of the ellipse, the corresponding relative percentage error in the radius was found to be between 0.5 and

**Table 7** The estimated radius  $R_{\text{est}}^f$  for 10 tomatoes is compared with the reference distances  $D_1$  and  $D_2$ . The distances are expressed in centimeters

S	12	13	14	15	16	17	18	19	20	21
$D_1$	2.53	2.07	2.09	2.51	2.48	1.44	1.89	2.57	2.13	2
$D_2$	2.03	1.92	1.75	2.2	2.14	1.36	1.51	2.14	1.86	1.77
$R_{\text{est}}^f$	2.35	1.99	2.02	2.58	2.36	1.55	1.86	1.86	1.96	1.85
$PE_{D_1}$	<i>7.20</i>	<i>3.57</i>	<i>3.05</i>	<i>2.46</i>	<i>4.76</i>	<i>7.63</i>	<i>1.54</i>	<i>27.79</i>	<i>7.79</i>	<i>7.43</i>
$PE_{D_2}$	15.64	3.97	15.50	17.40	10.14	13.96	23.23	12.91	5.34	4.59

The error percentage  $PE_{D_1}$  less than 10 % is shown in italics





**Fig. 19** For the sequence 19, the position of the tomato is not correctly detected. The manual segmentation is shown in red while  $Ell_{opt}$  is shown in green

3 % (0.6 and 3 %, respectively) depending on the position of the object in the scene, which is acceptable for the considered application.

## 7 Conclusions

This paper presents a complete system to monitor the growth of tomatoes from images captured in open field. One of the major challenges is occlusion. Moreover, poor illumination and the presence of neighboring tomatoes may cause the tomato contours to be smoothed, resulting in imprecision on their actual position. To overcome these challenges, we proposed to model the tomato as a sphere in the 3D space. This enables us to introduce a priori shape information in the segmentation procedure, which increases the robustness with respect to occlusion and lack of contrast. Besides, the spherical hypothesis allows us to simplify the size estimation procedure.

The segmentation method presented in this paper is an extension of our previous work [7, 8] based on active contour. In this paper, we propose to estimate the movement of the tomato between two consecutive images, by comparing SIFT descriptors computed at points of the contour. This leads to a more accurate estimate of the position of the tomato than with the pattern matching approach presented in [7, 8]. The improvement is more prominent in the images with significant occlusion (between 30 and 50 %) and poor contrast. For instance, the descriptor-based approach correctly detects the position in 96.5 % of the images where occlusion is between 30 and 50 %, which

is an improvement of +5 % compared to the previous approach. Moreover, a high accuracy is reached for 85 % of these images against 59 % with the previous approach. The precision is also significantly improved for low occluded images (97 % against 65 %). This is very important since a more accurate estimation of the tomato position results in a more reliable estimation of the candidate contour points, which in turn leads to a better initialization of the elliptic active contour model. So, the entire segmentation procedure benefits from this new algorithm. The average error (expressed as the percentage of tomato size) is now around 4 % even for tomatoes with a degree of occlusion as high as 50 %.

We also presented a method to estimate the size of the tomato from the obtained segmentation. This method was first tested under ideal acquisition conditions and using manual segmentation. In this case, the percentage error between the actual radius and the estimated size was always less than 10 % with most (91 %) of the error less than 5 %, which demonstrate the robustness of radius estimation. The complete system was also applied to estimate the size of tomatoes cultivated in open fields for the agriculture season 2013. The percentage error was less than 10 % in most of the cases, despite the poor quality of images during this season (small size, pixelated images).

The segmentation procedure based on shape information in each image separately can be extended to include the information in both images in order to propose a joint energy minimization scheme. For instance, if we suppose that a 3D space point  $X$  situated on the tomato is projected onto  $x_L$  in the left image and  $x_R$  in the right image; then, the evolution of the contour in the two images can be controlled by using the epipolar constraint ( $x_L^T F x_R = 0$ ) in a joint energy minimization functional, where  $F$  is the fundamental matrix computed from the two camera matrices. This approach would increase the robustness of the segmentation procedure with respect to occlusion, particularly in the image pairs where the percentage of occlusion is not identical.

One of the possible approaches to improve the robustness of the yield estimation would be to detect automatically the number of tomatoes present in an image without necessarily performing the segmentation procedure. This could be done during the end of the season when most tomatoes are red. By exploiting the color information, the density of tomatoes could be determined and combined with the size estimation performed on a subset of tomatoes, acquired with higher image resolution. This strategy would result in a more accurate estimate of the yield before the harvest. Moreover, the correction factor  $\alpha_{acc}$  involved in volume computation was estimated using a small set of tomatoes (Section 5.4). Further studies need to be done to develop a more accurate volume estimation model. The first experimental results obtained during



the agricultural season 2013 were very encouraging. However, we plan to conduct larger experiments in open fields to assess the robustness and the accuracy of the entire system.

In the future, we wish to integrate the proposed algorithm in a gateway/platform-based machine to machine (M2M) architecture in order to develop an operational system for the farmer to remotely monitor the growth of tomatoes. The proposed system may also be used to monitor the growth of other crops such as apples.

#### Competing interests

The authors declare that they have no competing interests.

#### Acknowledgements

This work was partly supported by the MCUBE project (European Regional Development Fund (ERDF)), which aims at integrating multimedia processing capabilities in a classical machine to machine (M2M) framework, thus allowing the user to remotely monitor an agricultural field. The authors would like to thank Jérôme Grangier, for his participation in this project. This work was performed while the first author was doing his PhD at ISEP and Telecom ParisTech.

#### Author details

<sup>1</sup>Department of Electronics and Communication Engineering, Manipal Institute of Technology, Manipal University, Manipal, India. <sup>2</sup>Institut Supérieur d'Electronique de Paris ISEP, Paris, France. <sup>3</sup>LTCL, CNRS, Télécom ParisTech, Université Paris-Saclay, Paris, France.

Received: 14 May 2015 Accepted: 20 October 2015

Published online: 04 November 2015

#### References

1. Estimating crop yields; a brief guide (2013). <http://agriculture.vic.gov.au/agriculture/grains-and-other-crops/crop-production/estimating-crop-yields-a-brief-guide>. Accessed October 2015
2. A Prasad, L Chai, R Singh, M Kafatos, Crop yield estimation model for Iowa using remote sensing and surface parameters. *Int. J. Appl. Earth Observation Geoinformation*. **8**(1), 26–33 (2006). doi:10.1016/j.jag.2005.06.002
3. M Mkhabela, P Bullock, S Raj, S Wang, Y Yang, Crop yield forecasting on the Canadian Prairies using MODIS NDVI data. *Agric. Forest Meteorol.* **151**(3), 385–393 (2011). doi:10.1016/j.agrformet.2010.11.012
4. H Zhao, Z Pei, in *second international conference on agro-geoinformatics (Agro-Geoinformatics)*. Crop growth monitoring by integration of time series remote sensing imagery and the WOFOST model, Fairfax, VA (IEEE, 2013), pp. 568–571. doi:10.1109/Argo-Geoinformatics.2013.6621940
5. D Stajanko, Z Cmelik, Modelling of apple fruit growth by application of image analysis. *Agric. Conspec. Sci.* **70**, 59–64 (2005)
6. A Aggelopoulou, D Bochtis, S Fountas, K Swain, T Gemtos, G Nanos, Yield prediction in apple orchards based on image processing. *J. Precision Agric.* **12**, 448–456 (2011)
7. U Verma, F Rossant, I Bloch, J Orensanz, D Boissongontier, in *international conference on pattern recognition applications and methods (ICPRAM)*. Shape-based segmentation of tomatoes for agriculture monitoring (Angers, France, 2014), pp. 402–411
8. U Verma, F Rossant, I Bloch, J Orensanz, D Boissongontier, *Segmentation of tomatoes in open field images with shape and temporal constraints, Pattern Recognition, Applications and Methods*. (A Fred, et al., eds.) (LNCS 9443: ICPRAM 2014 Best Papers, Springer, 2015). (forthcoming)
9. J Bouguet, Camera calibration toolbox for Matlab (2013). [http://www.vision.caltech.edu/bouguetj/calib\\_doc/](http://www.vision.caltech.edu/bouguetj/calib_doc/). Accessed April 2013
10. R Hartley, A Zisserman, *Multiple View Geometry in Computer Vision*, 2nd edn. (Cambridge University Press, New York, NY, USA, 2004)
11. M Kass, A Witkin, D Terzopoulos, Snakes: active contour models. *Int. J. Comput. Vis.* **1**(4), 321–331 (1988)

12. C Xu, J Prince, Snakes, shapes, and gradient vector flow. *IEEE Trans. Image Process.* **7**(3), 359–369 (1998)
13. D Lowe, Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **60**(2), 91–110 (2004)
14. W Gander, G Golub, R Strebel, Least-squares fitting of circles and ellipses. *BIT Numerical Math.* **34**(4), 558–578 (1994)

**Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:**

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](http://springeropen.com)