Towards Autonomous Detection and Tracking of Electric Towers for Aerial Power Line Inspection

Carol Martinez, Carlos Sampedro, Aneesh Chauhan, and Pascual Campoy

Abstract— This paper presents an approach towards autonomous aerial power line inspection. In particular, the presented work focuses on real-time autonomous detection, localization and tracking of electric towers. A strategy which combines classic computer vision and machine learning techniques is proposed. A generalized detection and localization approach is presented, where a two-class multilayer perceptron (MLP) was trained for Tower-Background classification. This MLP is applied over sliding windows for each camera frame until a tower is detected. The detection of a tower triggers the tracker. A hierarchical tracking methodology, especially designed for tracking towers in real-time, is presented. This methodology is based on the Hierarchical Multi-Parametric and Multi-Resolution Inverse Compositional Algorithm [1], and is proposed to be used for tracking and maintaining the tower in the field of view (FOV). The proposed strategy, which is the combination of the tower detector and the tracker, is evaluated on videos from several real manned helicopter inspections. Overall, the results show that the proposed strategy performs very well at detecting and tracking various types of electric towers in diverse environmental settings.

I. INTRODUCTION

Electricity companies have several regulatory as well as financial reasons for performing regular inspections of the power line infrastructure [2], [3]. The faults and defects in the power line equipment are a big cause of unplanned power outages, as well as present a public safety risk [2], [3]. Consequently, these companies spend a significant budget on the power line inspections, and continuously pursue new approaches to reduce the inspection costs.

These issues have led to several collaborative efforts between industry and academia to research other approaches, with focus on automating the inspection task. From the field of mobile robotics, two approaches have primarily emerged [4], [5]: Unmanned Aerial Vehicles (UAVs) [6], [7], [8], [9], and the robots capable of climbing and rolling on the wires (ROW) [10], [11], [12]. Some authors have also proposed hybrid climbing-flying robots [13].

Both types of robotic platforms offer promising solutions. However, UAVs have one major advantage over ROWs [4], [5]: they offer a general solution to inspection, that is, unlike ROW robots, it is not necessary to adapt the robot design to different types of towers and power line infrastructure. UAVs offer a very attractive solution for automating the task of power line inspection. Having said this, to provide

Authors are with the Computer Vision Group (www.vision4uav.com), Centro de Automática y Robótica (CAR) UPM-CSIC, Universidad Politécnica de Madrid, José Gutiérrez Abascal 2, 28006 Madrid, Spain. Email: {carolviviana.martinez, pascual.campoy}@upm.es

a completely autonomous UAV-based inspection, several challenges need to be addressed.

In a recent paper, Pagnano et al [5] highlighted some of the most important general challenges for UAV-based power line inspection:

- Visual servoing (extended by information from other sensors) for power line tracking and autonomous navigation;
- Obstacle detection and avoidance (a UAV should not crash into the power line equipment);
- Robust control (for high stability and positioning for close-up and comprehensive inspections).

In resolving these general challenges, computer vision plays a major role. Different aspects of UAV-based power line inspection can benefit greatly from appropriate vision algorithms.

This paper is a contribution towards autonomous aerial inspection of the power line infrastructure. More specifically, the problem of autonomous detection and tracking of electric towers is addressed based on classic computer vision and machine learning techniques.

Several researchers have applied computer vision techniques for tower detection and segmentation in aerial images [14], [15], [16], [17], [18]. Since towers are usually linear structures, most approaches are based on detecting lines in an image. The detected lines are post-processed by applying user-defined heuristics, in order to keep only the lines belonging to the tower. Various image segmentation methods are then applied to extract the tower from the image - e.g. direct template match is used in [15], watershed segmentation in [16], graph-cut in [17]. Golightly and Jones [14], instead of using lines, used corner features to detect a tower in the image.

The problem of detection and tracking of towers during aerial inspections has been addressed previously for remotely piloted helicopter missions [15], [14]. Whitworth et al [15] apply a template-based approach for tower detection and tracking. The template matching approach was designed for detecting and segmenting simple "T-shaped" towers from the video sequences. The same approach is then recursively applied for tracking as well. Golightly and Jones [14] use a corner detector to detect and track the tower-tops where the detected corners exemplify a tower-top in an image. The results were reported only for a single type of tower.

A major restriction of these approaches, for detection as well as tracking, is that they have been reported on just one type of tower, and authors make several assumptions regarding the shape and appearance of the tower. Moreover,

the robustness of the proposed approaches have not been tested on diverse backgrounds, as would be expected from inspections. Thus, current approaches to tower detection and tracking cannot be generalized. Fig. 1 shows some examples of different types of electric towers and heterogeneity in background, as observed in real aerial inspections.

In this paper, a general approach to tower detection is presented. A multilayer perceptron (MLP) neural network was trained for Tower-Background classification. The detector was trained on the cropped images of four types of towers and background regions. HOG (Histogram of Oriented Gradients) features [19] are extracted from these images and used as input for training the MLP. To detect the tower in an image, the sliding window approach (as in [20], [21]) is used where the trained MLP is applied to image regions. A successful detection of tower in an image triggers the tower tracking routine.

In this paper, the Hierarchical Multi-Parametric and Multi-Resolution Inverse Compositional Algorithm (HMPMR-ICIA) [1], a tracking-by-registration strategy, is explored for tower tracking. The intensity information of all the pixels of the electric tower (or part of the tower) are used to estimate the motion of the tower in the image plane. Direct methods permit to obtain robust estimations and also have the advantage of being more generally applicable to a wide range of scenes (selection and extraction of specific features are not required). The latter advantage is of great importance for the application concerned in this paper where tracking towers with different kind of sizes and shapes is required, as well as obtaining robust estimations under the difficult conditions of the task (constant vibrations of the UAV, high-speed flights, and the limited computational capacity on-board). This is why the HMPMR-ICIA algorithm is used.

Therefore, the main contribution of this work is a general vision-based strategy for tower detection and tracking. Unlike other works related to the use of computer vision for power line inspection, the proposed strategy does not make any assumption on the shape of the tower to be detected and tracked. Additionally, most of the approaches found in the literature do not take tracking into account. To the authors' knowledge, this is the first work where a robust machine learning-based tower detection approach is combined with a tracking-by-registration approach to solve the tower detection and tracking problem for autonomous power line inspection.

The rest of the paper is structured as follows: Section II presents the problem statement; Section III describes the proposed strategy for tower detection and tracking; in Section IV, the proposed strategy is evaluated using image data from two real manned helicopter inspections; and Section V concludes the paper.

II. PROBLEM STATEMENT

Electric towers are complex structures with several critical components that are prone to damage and deterioration. Therefore, a comprehensive visual inspection of the towers is an extremely important part of the power line inspection. A UAV platform with on-board visual sensing and processing

Fig. 1. Power line inspection. The inspection of power line infrastructures is a very challenging task, for computer vision algorithms, due to the large variety of backgrounds and types of tower.

equipment can greatly facilitate the autonomous inspection task.

As elaborated in the previous section, computer vision techniques are widely applied for processing the aerial images for tower detection and tracking. Further image processing algorithms can then be applied for detecting the power line components on the tower (joints, conductors, insulators, etc.) and for analyzing the faults that can appear on them. Nevertheless, the application of computer vision for power line inspection is a very challenging task.

One of the problems that computer vision has to deal with, is the wide variety of backgrounds in which the images are taken. Backgrounds can be very heterogeneous in color, texture, and also in the elements that can appear in the scene, like houses, roads, crops, etc. (see Fig. 1), which can make the inspection task difficult. Illumination changes are yet another factor that adversely impact the performance of the algorithms. Due to this, it is very difficult to make general assumptions about the color of the tower or the background for processing the image.

Other important challenge for computer vision techniques is the wide variety of shapes and sizes that electric towers can have, e.g., electric towers used for medium-voltage power lines (see bottom-right image in Fig. 1) have heights about 10m, and are mainly composed of a vertical pole with the insulators placed on the top of the tower. On the other hand, electric towers used for very high-voltage power lines (see top-images in Fig. 1) can reach up to 55m, and have a more complex structure.

Another important factor is the quality of images. It depends considerably on the kind of inspection. In intensive aerial inspections, the flights are conducted at very low speed and stop at every tower, pointing the camera at the components to be inspected. In such inspections the image resolution is high, and therefore, small-sized faults in the components can be better detected and analyzed. In non-

intensive inspections, the flight is conducted at high speed, making a continuous flight without stopping at each tower. In this case, the images are of poorer quality, and only some faults in the structure of the electric towers can be detected. These problems appear with manned helicopter as well as UAV-based inspections.

In order to have an operating UAV-based platform completely autonomous, it is important to have a tracking approach working on-line for image acquisition and processing. This will help the UAV to navigate in order to maintain the tower and its components inside the filed of view of the camera as soon as the tower is detected. Furthermore, the detection techniques are usually computationally more expensive than tracking methods. Therefore, for designing a real-time online solution, it would be useful to avoid tower detection in every single frame, and instead apply tracking. The tracking algorithm will also have to deal with several problems like abrupt movements of the camera, scale changes of the electric tower and its components. Additionally, there is a clear interdependency between the tracker and the detector, since the detector output provides the region for the tracker to track.

Currently there is no global system for automating the power line inspection process that can be generalized for all the tower types (medium, high, very-high voltage, etc) and in all the possible scenarios. Although simultaneous tower detection and tracking has been previously addressed in [15], [14], their experimental scenario was limited to a single type of tower in a simpler environmental setting.

Considering all the mentioned restrictions, this paper is conceived as an effort to detect and track electric towers in a large variety of scenarios which can appear in images acquired by real non-intensive inspections. The main motivation of the approach presented here is to help the automation of the power line inspection process.

III. TOWER DETECTION AND TRACKING STRATEGY

In this paper, a tower detection and tracking strategy is explored with the aim of automating the power line inspection problem using Unmanned Aerial Vehicles (UAVs). The objective of the system is first to determine the position of the tower in the image plane using a machine learning approach, and then estimating the position of the tower in the following frames using a tracking-by-registration approach.

A. Strategy Overview

Fig. 2 presents the flowchart of the tower detection and tracking strategy. As can be seen in this figure, at the core of the system are the tower detection and the tower tracking stages. The algorithm starts with status $TD = 0$, (tower detection flag) i.e. electric tower is not detected. The tower detection stage, described in Section III-B, is then used to find an electric tower in the current frame (e.g. in $\mathbf{I}_{(0)}$, where the subindex represents the number of the frame). Detection criteria, described in Section III-D, are then applied in order to define if the found region contains a tower. If those criteria
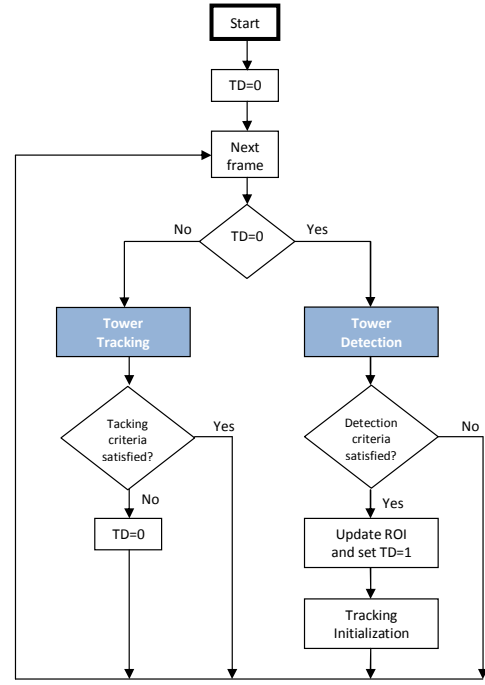

Fig. 2. System architecture. The proposed strategy for power line inspection is based on the interaction between a detection and a tracking stage. Tracking and detection criteria contra interaction between both stages, based on the TD flag (Tower Detection flag).

are met, $TD = 1$, and the found region is post-processed in order to obtain a more compact region. This updated region will be the template image $\mathbf{T}_{(0)}$ (sub-index refers to the image number) used by the tracking-by-registration algorithm (described in Section III-C). Then, the tracking algorithm is initialized. Different components of the tracking-by-registration strategy, such as the pyramidal structure, the Hessian matrix, etc., are created and calculated. As can be seen in Fig. 2, the tracking initialization stage is carried out every time the template image is updated.

When a new frame is analyzed, if $TD = 1$, then the tracking algorithm is used to estimate the transformation that describes the motion of the electric tower from the frame where $\mathbf{T}_0$ was selected, to the current frame.

Then, as can be seen in Fig. 2, the result of the tracking algorithm is checked by different criteria which analyze either the performance of the tracking algorithm or the position of the tower in the image plane (e.g. if the tower is too close to the camera, the algorithm triggers $TD = 0$, in order to search the next electric tower).

These tracking criteria are used to switch between the detection and tracking stages. If some of those criteria are not satisfied, tower detection flag is switch to $TD = 0$, and the detection stage will operate until a new electric tower is found. Conversely, if the criteria are satisfied, the position of the tower in the current frame is known.

Most of the time, the tracking stage operates in isolation because this is the most robust and computationally efficient approach to determine the position of the tower in the different frames of an image sequence. Nevertheless, the tower

Input Image   Vertical Sobel   Hough Transform   Define Search Area

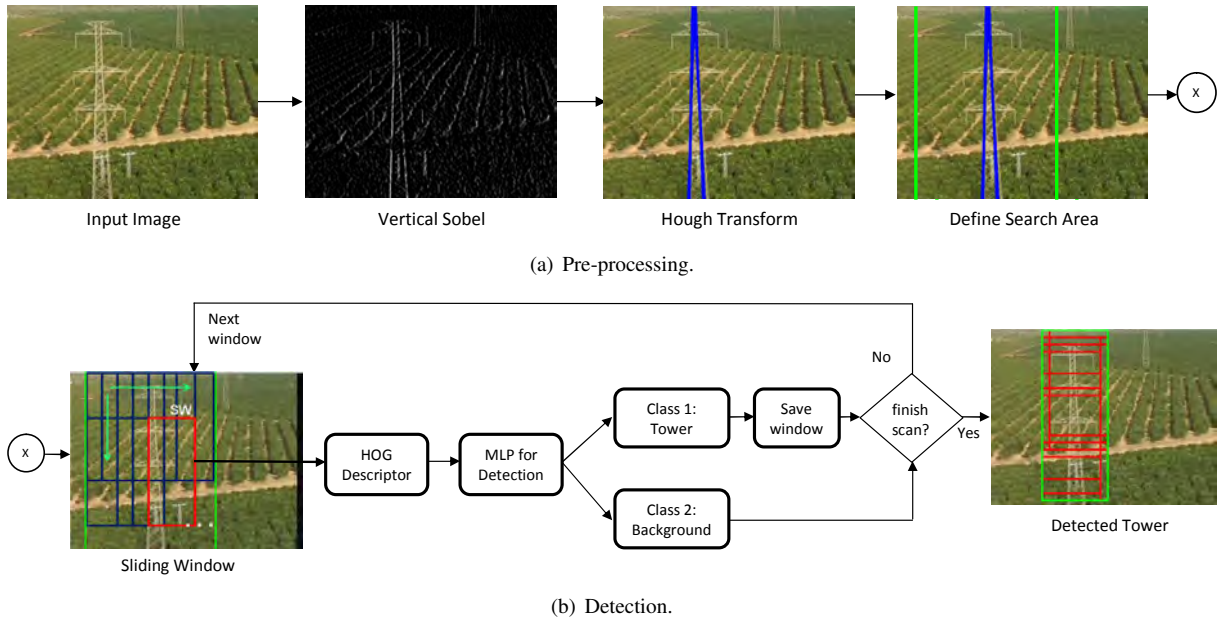(a) Pre-processing.

(b) Detection.

Fig. 3. Tower detection stage. Given an input image, a preprocessing step is applied to detect vertical lines. Once the lines are computed, a search area for the sliding window is defined. The sliding window approach is used to scan the search area. A trained tower detection classifier is applied in each window (**SW**) for Tower-Background classification.

detection stage acts as a backup to detect the position of the tower whenever the tracking stage needs to be reinitialized.

B. Tower Detection Stage

This stage is in charge of detecting and localizing electric towers in an image. As can be seen in Fig. 2, if some criteria are satisfied, the output of the detection stage is used to initialize the tracking algorithm.

Fig. 3 describes the workflow of the tower detection stage. This stage is composed of two algorithms, one based on detecting vertical lines (Fig. 3(a)) and a second one based on tower-background classification (Fig. 3(b)). The general idea of the strategy is to use a sliding window approach to scan the image and to apply the trained tower detection classifier in each window. Nonetheless, in order to speed up the process and to improve the precision of the tower detection algorithm, the sliding window approach is applied only in the regions that contain vertical lines which are defined in a pre-processing stage described in Fig. 3(a).

In the pre-processing algorithm (see Fig. 3(a)), the vertical Sobel mask is applied over the input image (converted to grayscale) in order to extract vertical edges. Then, the vertical sobel image is thresholded in order to reduce noise. This threshold has been found experimentally ($sThr = 40/255$) finding a compromise between noise reduction and vertical edge detection. The Hough transform is then applied over the thresholded image to detect vertical lines. It is applied only in the range of $-5°$ to $5°$, in order to reduce processing time. Then, a threshold in the hough space ($hThr = 200$) is applied in order to keep the lines that have a length bigger than a defined value. The length is controlled by the number of votes of each line in the Hough space which have been defined experimentally. Therefore, those lines with a number

of votes $> hThr$ are considered as possible lines belonging to towers. Finally, the position of the lines found in the image plane is used to define the search area for the sliding window algorithm, as can be seen in Fig. 3(a). The size of the search area is based on the size of the windows used by the sliding window algorithm.

In order to detect the electric tower, a sliding window approach is used to scan the search area defined in the pre-processing stage. As shown in Fig. 3(b), a small window, **SW**, of a predefined size is slid over the image. In our strategy two window sizes are used (**SW**1:160×290 pixels, and **SW**2:130×260 pixels). The size of these windows is derived from the average size of the tower images used for training the classifier.

Each window **SW**, provided by the sliding window algorithm, is resized to 64×128 pixels, and then from this image HOG (Histogram of Oriented Gradients) features [19] are extracted. The resulting HOG feature vector (of size 3780) is passed as input to a two-class MLP classifier trained for tower detection. The window **SW** will be classified as Tower, if the activation values of the output layer neurons, a_1 and a_2, are $a_1 \geq 0.98$ & $a_2 \leq 0.001$; or as Background, otherwise.

The position in the image of all the windows **SW** that have been classified as Tower are then saved (see Fig. 3(b), red boxes, image on the right). Finally, when the sliding algorithm has finished scanning the entire image, the result from the detection stage is obtained as the bounding box that covers all windows **SW** that were saved. This region of interest, shown in Fig. 3(b) (green box, image on the right), is post-processed in order to obtain a more compact region. The updated region $\mathbf{ROI}_{(F)}$, is obtained by extracting vertical lines inside the region found by the detection stage. Thus,

$\mathbf{ROI}_{(F)}$ corresponds to the rectangle that inscribes the vertical lines of the tower. $\mathbf{ROI}_{(F)}$ will be used as the image template, $\mathbf{T}_{(F)}$, for the tracking stage.

C. Tower Tracking

The selected strategy for tracking electric towers is the HMPMR-ICIA (Hierarchical Multi-Parametric and Multi-Resolution Inverse Compositional Algorithm) proposed in [1]. This is a tracking-by-registration algorithm based on direct methods, which have shown to have fast and robust performance for tracking objects with cameras on-board aerial vehicles.

In this paper we explore the use of this technique for tracking electric towers. The advantages of using direct methods is that they make an optimal use of the available information in the image (using the intensity values of all the pixels of a defined region in order to estimate the motion), therefore, direct methods permit to obtain robust estimations and also they have the advantage of being more generally applicable to a wide range of scenes. The latter advantage is of great importance for the application concerned in this paper in which tracking towers with different kind of sizes and shapes is required.

The goal of the tracking algorithm is to find the 2D position of the tower in each frame (F) of an image sequence (i.e. $\mathbf{I}_{(F)}$). The following reasonable assumptions were made in the proposed strategy: the 2D position is found assuming that an initial position of the tower in the previous frame is known, the 3D motion of the tower projected in the image plane can be modeled by a 2D transformation, and assuming the direct methods constraints [22] (small frame-to-frame motion, brightness constancy constraint, etc.).

An example of the general idea of the tracking-by-registration algorithm for tower tracking can be seen in Fig. 4. A reference image ($\mathbf{T}_{(0)}$) is defined in the first frame (Fig. 4 Frame 0, upper left image). This reference image corresponds to a sub-image or $\mathbf{ROI}_{(0)}$ (Region of Interest), called image template, that contains a tower. This $\mathbf{ROI}_{(0)}$ is found by the detection stage, as explained in Section III-B

When a new frame is analyzed e.g. $\mathbf{I}_{(1)}$ (Fig. 4, Frame 1, upper right image), the motion $\mathbf{W}_{(1)}$ between $\mathbf{I}_{(0)}$ and $\mathbf{I}_{(1)}$ (Fig. 4, Frame 1, green solid arrow) is found by an image registration technique, assuming that an initial estimation of the motion $\mathbf{W}_{\mathrm{init}}$ is known (Fig. 4, Frame 1, yellow/dashed arrow). Thus iteratively estimating the incremental motion model $\mathbf{\Delta W}$. When an initial estimation is not known, it can be assumed as the identity matrix when the frame-to-frame motion is small. Therefore, the motion $\mathbf{W}_{(1)}$ is estimated by the image registration technique, and as a consequence of this, $\mathbf{ROI}_{(1)}$ is found, i.e. the position of the tower in current frame (e.g. $\mathbf{I}_{(1)}$).

The estimated motion $\mathbf{W}_{(1)}$ (Figure 4, Frame 1, green/solid arrow) is propagated to the next frame, as an initial estimation of the motion $\mathbf{W}_{\mathrm{init}} = \mathbf{W}_{(1)}$ (yellow/dashed arrow, Frame 2, bottom left image). The process is repeated with each frame of the sequence and therefore, the tower is tracked throughout the sequence.
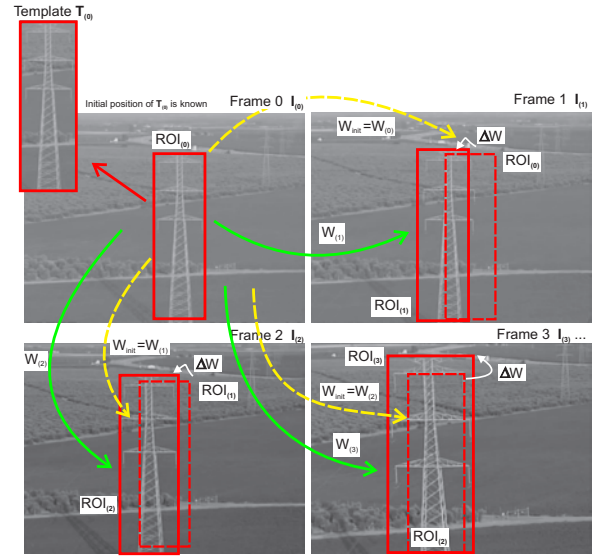

Fig. 4. An example of tracking-by-registration strategy. The ROI that contains a tower is defined in the first frame $\mathbf{ROI}_{(0)}$. It is found by a tower detection algorithm. When a new frame is analyzed, e.g. $\mathbf{I}_{(1)}$ (upper right image), the motion $\mathbf{W}_{(1)}$ between $\mathbf{I}_{(0)}$ and $\mathbf{I}_{(1)}$ (green/solid arrow) is found by an image registration technique, assuming that an initial estimation of the motion $\mathbf{W}_{\mathrm{init}}$ is known (yellow/dashed arrow).

1) Motion Model: The transformation or motion model $\mathbf{W}$ that maps the pixels $\mathbf{x} = (x, y, 1)^{\mathrm{T}}$ from $\mathbf{T}_{(0)}$ to pixels $\mathbf{x}' = (x', y', 1)^{\mathrm{T}}$ in frame $\mathbf{I}_{(F)}$ is defined as follows:

$$\mathbf{x}' = \mathbf{W}\,\mathbf{x} = \mathbf{W}(\mathbf{x}; \mathbf{p})$$

$$\mathbf{W} = \begin{bmatrix} 1 + p_1 & p_2 & p_3 \\ p_4 & 1 + p_5 & p_6 \\ p_7 & p_8 & 1 \end{bmatrix} \quad (1)$$

where this motion model $\mathbf{W}$ is parameterized by the vector of parameters $\mathbf{p} = (p_1, ... p_n)^{\mathrm{T}}$ in such a way that $\mathbf{W}$ is the identity matrix when the parameters are equal to zero. Depending on the values of $\mathbf{p}$, $\mathbf{W}$ can model different 2D transformations, e.g. if p_1, p_2, p_4, p_5, p_7 and p_8 are equal to zero, then the translation motion model is estimated (parameterized by p_3 and p_6).

2) Tracking-by-Registration: As mentioned in Section III-C, at the core of the tracking-by-registration strategy is an image registration technique. In this paper, an image registration algorithm based on direct methods is used. The image registration algorithm used to estimate the parameters of $\mathbf{W}_{(F)}$ is the ICIA algorithm [23] extended with a HMPMR strategy [1] in order to deal with the small motion constraint of direct methods and with the efficiency problem of these methods.

The HMPMR-ICIA algorithm makes use of two hierarchical structures: the Multi-Resolution (MR) and the Multi-Parametric (MP) ones. The general idea behind this strategy is that by estimating only a small number of parameters at the lowest resolution levels and smoothly increasing the complexity of the motion model through the MR pyramid, it is possible to obtain a more robust and a faster estimation

of motion models under large frame-to-frame motions than the one obtained when only a MR approach is used.

The two hierarchical structures of the HMPMR strategy are created as follows. The MR structure is created by repeatedly downsampling the images by a factor of 2 [24] in order to create the different pyramid levels (pL). In the MP structure, the number of parameters increases with the resolution of the image. The MP structure is defined according to the motion model selected at the lowest level of the pyramid (the highest resolution level) $\mathbf{W}_{(F)}^{0}$ (where the superscript represents the level of the pyramid and the subscript the frame). In this level, $\mathbf{W}_{(F)}^{0}$ must be chosen as the best transformation that represents the motion of the object in the image plane. Additionally, in order to ensure the detection of large frame-to-frame motions, the translation motion model must be chosen for the highest level of the pyramid (the level that has the lowest resolution image) $\mathbf{W}_{(F)}^{j_{\max}}$ (where $j_{\max} = pL-1$). Finally, the criterion to define the parameters in the intermediate levels is to select them in such a way that a smooth transition of the parameters from the highest to the lowest level of the pyramid is obtained, as described in [1].

The HMPMR-ICIA minimizes (using a gradient descent approach), in each level j (where, $j = \{pL - 1, pL - 2, \ldots, 0\}$), the Sum of Squared Differences (SSD) between the image template $\mathbf{T}_{(0)}^{j}$ and the current image $\mathbf{I}_{(F)}^{j}$:

$$\sum_{\mathbf{x}} [T_{(0)}^{j}(\mathbf{W}_{(F)}^{j}(\mathbf{x}; \Delta\mathbf{p}^{j})) - I_{(F)}^{j}(\mathbf{W}_{(F)}^{j}(\mathbf{x}; \mathbf{p}^{j}))]^{2} \quad (2)$$

Where $T_{(0)}^{j}$ is the template image found by the tower detection algorithm (see Section III-B), scaled according to each multi-resolution level j; $I_{(F)}^{j}$ is the scaled current image at level j; $\mathbf{x} = (x, y)^{\mathrm{T}}$ represents the pixel coordinates; and $\mathbf{W}_{(F)}^{j}(\mathbf{x}; \mathbf{p})$ is the motion model that will be estimated in each level j, where $\mathbf{p}^{j} = (p_{1}, p_{2}.., p_{i})^{\mathrm{T}}$ is the vector of parameters that describes the transformation in each level.

As mentioned in Section III-A, some information required by the HMPMR-ICIA algorithm is calculated only once in the initialization stage and every time the template image is updated. Some of these parameters are the number of levels of the hierarchical structure, the MP structure, the MR structure of $\mathbf{T}$, the Hessian matrix, amongst others [1].

The HMPMR-ICIA algorithm iteratively updates the parameters of the motion model, in each level of the pyramid, until stopping criteria are reached denoting the best local alignment solution. Therefore, at the lowest level of the pyramid (i.e highest resolution image), the most complex motion model is estimated. This motion model is the best approximation of the motion of the object in the image plane. With this information, it is possible to determine the position of $\mathbf{T}_{(0)}$ in the current image $\mathbf{I}_{(F)}$.

An important part of the HMPMR structure is the propagation of parameters inside the MR structure and among frames. The parameters that are estimated in each level are used as an initial estimation of the motion for the following levels. Additionally, in level 0, another kind of propagation

occurs from the lowest level of the pyramid of the previous frame $\mathbf{I}_{(F-1)}$, to the highest level of the pyramid ($j = j_{max}$) of the new frame $\mathbf{I}_{(F)}$. This propagation is the basis of the tracking-by-registration strategy depicted in Fig. 4.

D. Switching Criteria

As can be seen in Fig. 2, in order to switch between the detection and tracking stages, performance assessment criteria were applied to monitor the behavior of these stages. The following criteria were used:

- Vertical lines length $> lThr$: if the length of the lines is larger than a threshold. This condition permits to determine if there are vertical lines present in an image.
- Vertical lines slope: if the slope of the vertical lines lie within a pre-defined range.
- c_x and c_y, the coordinates of the center of the image $\mathbf{ROI}_{(F)}$: if c_x is not close to zero, or if c_y is not close to the image height.
- $\mathrm{MAE}^{j} > maeThr$: in the tracking stage, if the mean absolute error (MAE) at each pyramid level is greater than a threshold. Where $\mathrm{MAE}^{j_{max}}$ is defined as:

$$\mathrm{MAE}^{j} = \frac{\sum_{\mathbf{x}} |T_{(0)}^{j}(\mathbf{x}) - I_{(F)}^{j}(\mathbf{W}^{j}(\mathbf{x}; \mathbf{p}))|}{n^{j}} \quad (3)$$

where n^{j} is the total number of pixels.
- $\Delta\mathbf{pos} < pThr$: the difference between the current and the previous position of $\mathbf{ROI}$ in the image plane is below a given threshold.
- % of pixels $> ppThr$: the percentage of pixels that are used in the minimization process is greater than a given threshold. This condition is particularly useful when the electric tower is leaving the FOV of the camera or when it is occluded.

The result of the tower detection stage is controlled by analyzing the existence of lines inside the found ROI. If the lines inside the ROI comply the vertical lines length criterion and the angle criterion, then the tower detection flag is switched to $\mathrm{TD} = 1$.

On the other hand, the result of the tower tracking stage is controlled by analyzing all the different criteria explained above. The main idea is to analyze the existence of lines inside the found ROI, to analyze the motion and position of the ROI, and to analyze some parameters of the tracking algorithm, such as the MAE and the number of pixels used to estimate the motion. If all the above mentioned criteria are not satisfied, then $\mathrm{TD} = 0$, i.e. the tracking algorithm is unable to determine the position of the tower with a high degree of confidence, and therefore it needs to be reinitialized.

All these criteria have been found experimentally, and are defined in Section IV-A.3.

IV. RESULTS

In this section, the performance of the proposed tower detection and tracking strategy is evaluated. Different tests have been conducted using videos from real inspections with different kinds of towers and backgrounds. Qualitative and quantitative results are presented.
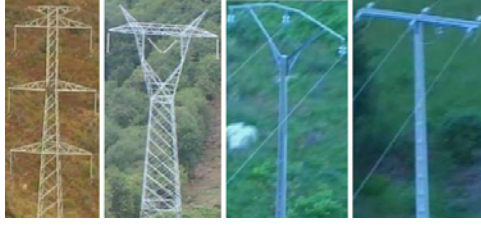
(a) Tower examples used for training.
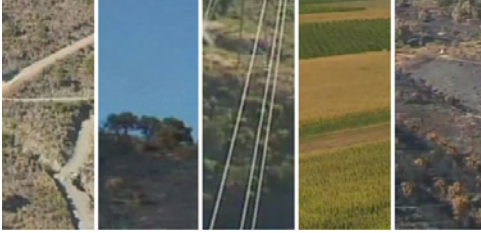
(b) Background examples used for training.

Fig. 5. Examples of cropped images of: (a) 4 cropped tower images and (b) 5 background images. These and similar images are used for training and evaluation of the MLPs for tower detection and tower-type classification.

A. Experimental Setup

1) Data Collection: Currently there are no publicly available datasets of power line inspection, and proprietary aerial inspection data was made available by an electric power company. The data consists of 11 videos captured during multiple non-intensive manned aerial inspections (helicopter was flying at $\approx$ 30 km/h). Six of these videos primarily contain inspections of towers supporting high voltage lines, such as the first and second image shown in Fig. 5(a). The other 5 videos contain inspections of towers for medium voltage lines (e.g. the third and fourth image shown in Fig. 5(a)).

From these videos, a dataset of cropped images was created in order to train the neural network used in the tower detection stage. These cropped images were either labeled as Background or as Tower. In total, 3200 image regions from 11 videos (1600 regions containing tower and 1600 containing background) were manually labelled (see Fig. 5 for sample images).

Additionally, two of the available videos were selected in order to test and evaluate the performance of the tower detection and tracking strategy. These videos contain three different types of electric towers: one that supports high voltage lines (first image in Fig. 5(a)); and two that supports medium voltage lines (third and fourth images in Fig. 5(a)).

2) Training and Evaluation Methodology of the MLP Classifier: The tower detection stage is based on an MLP classifier. In order to train and evaluate the MLP, 3200 images were divided into training (1200 images per class), cross validation (200 images per class), and test sets (200 images per class). It should be noted that the images belonging to the Tower class come from 4 types of towers. Therefore, 300 images of each type are used for training, 50 each for

validation and 50 each for testing.

3) Strategy Setup: The tracking algorithm (HMPMR-ICIA) has been configured with 3 pyramid levels. The MP structure was configured as 4-2-2, where each number represents the number of parameters estimated in each level. In the lower resolution levels, the translation motion model is estimated (2 parameters), and in the highest resolution level, the similarity transformation [25] (translation, rotation, and scale) is estimated (4 parameters). This motion model has been selected taking into account that during a visual inspection, the UAV will fly next to the power line and the dominant motion of the tower is mainly described by the changes in position, in scale (the tower approaches the camera), and sometimes small rotations.

For minimizing the cost function (2), the following criteria were considered: in each level, the number of iterations is fixed at 100 iterations per level; and two termination criteria are applied: when the increment of the parameters is below a threshold (10^{-5}), or if the mean error does not decrease after a defined number of iterations (10 iterations).

On the other hand, when direct methods are used, all the pixels of the template image are used to estimate the motion of the object. For this reason, the template image used for tracking electric towers will correspond to an area defined around the vertical lines of the tower. With this criterion, the amount of background information used for motion estimation is reduced. This is especially useful for towers of high voltage lines. In the results presented in this paper, due to the low quality of the available images, vertical lines are not always well detected, e.g. in high voltage towers one line is sometimes detected instead of two main lines. For this reason, in the test, the type of tower is used as an input parameter, so that the proper size of the template image can be defined, in case not all the lines are well detected.

In the conducted experiments, the values of the different switching criteria, explained in Section III-D, were set to the following values: $lThrL = 150$; vertical line slope range $< 10°$ and $> 150°$; $c_x > 20$ and $c_y <$ image width; $\text{MAE}^0 > 25$ and $\text{MAE}^1 > 20$, $\text{MAE}^2 > 20$; $\Delta\text{ROI}_x, \Delta\text{ROI}_y < 20$ (upper left coordinates of the image **ROI**) and $\Delta\text{ROI}_w, \Delta\text{ROI}_h < 35$ (width and height of image **ROI**); and $ppThresh = 70$.

B. Evaluation of the Detection Stage

Table I shows the confusion matrix obtained when testing the MLP. A total test error of 3.25% is achieved. The false positive rate obtained is 2.5%, which means that only 5 images of the 200 background cropped images were misclassified as tower. This result is very encouraging, especially taking into account the large variety of backgrounds which appear in the images (Figure 5(b)). On the other hand, a false negative rate of 4% was obtained. This indicates that only 8 images of the 200 cropped images containing got misclassified as background. This result is also very promising, taking into account the variety of tower shapes and sizes in the dataset.

Predicted class \ Actual class	Tower	Background
Tower (%)	96	2.5
Background (%)	4	97.5

The proposed tower detection stage, described in Section III-B, has been evaluated using 110 new images (not cropped like the ones used in the MLP evaluation). In these images, 60 contained an electric tower (15 per each type of tower), and 50 contained only background. Table II(a) shows the results obtained when the tower detection stage, described in Fig. 3 was tested, using both the pre-processing and the detection algorithm. Additionally, the results applying only the detection algorithm (using only the sliding window approach), without pre-processing step are shown in Table II(b).

From these results, it can be observed that by applying a pre-processing step, based on line detection, all the false positives, that were obtained when the detection stage was applied without pre-processing algorithm, were removed. Therefore, the strategy proposed in this paper for tower detection, obtained a 0% false positive rate in the 110 image test set. That is, in all the images that contain only background, no tower were detected. On the other hand, the false negative rate was 10%, which indicates that only 7 of the 60 images that contain only electric towers, were misclassified or not detected.

(a) Result of the Detection Stage with preprocessing

Predicted class \ Actual class	Tower	Background
Tower (%)	90	0
Background (%)	10	100

(b) Result of the detection Stage without preprocessing

Predicted class \ Actual class	Tower	Background
Tower (%)	91.67	26
Background (%)	8.33	74

C. Evaluation of the Tower Detection and Tracking Strategy

The proposed strategy for tower detection and tracking has been evaluated using two image sequences from real power line inspections conducted with a manned helicopter. The sequences contain different kinds of towers and background, fast motions (scale and perspective changes), part of the tower outside the FOV of the camera, and strong motion blur, among other features, which make these sequences challenging from the detection and tracking point of view.

In both tests, the analysis of the results is based on a visual examination of the results (analyzing the ROI found by the algorithms); and also based on a comparison between the results obtained when only the detection stage is applied and the results obtained when the complete strategy was applied. The idea of the comparison is to highlight the advantages of combining the detection and the tracking approaches.

1) High Voltage Tower: The image sequence used in this test contains the inspection of 11 high voltage towers (7177 frames). Some of the main features of the images in this sequence are their low image quality, the different types of background (which make it difficult, even for a human, to see the tower), motion blur, and that part of the tower is out of sight (during the transition from tower to tower and when the tower is closer to the camera).

In Fig. 6, a collection of images illustrating the performance of the proposed tower detection and tracking strategy, is presented. The red/dark box indicates the results.
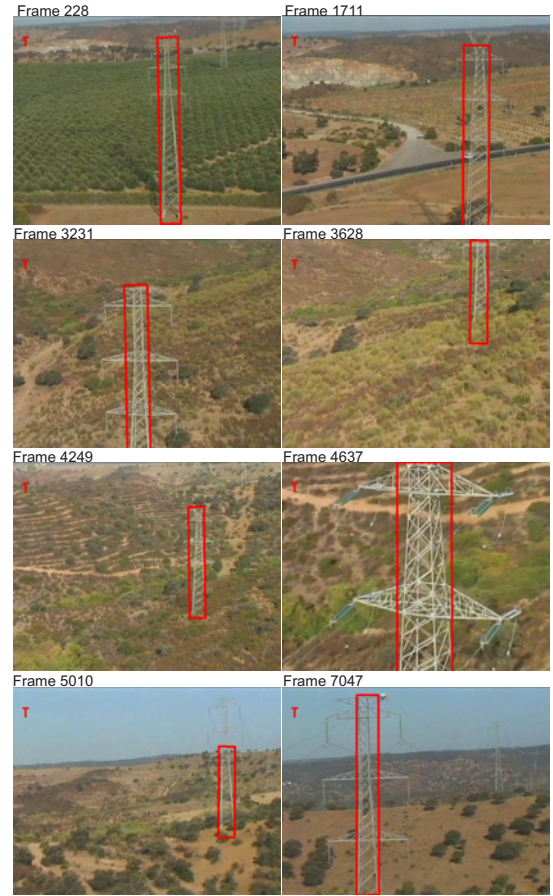

Fig. 6. Sample results of applying the complete detection and tracking strategy in an image sequence with high voltage towers. The red/dark box indicates the results.

As was mentioned in Section III-A, when the complete strategy is used the tower detection stage starts operating until a tower is found in an image. Then, the tracking stage gets in charge of tracking the region found by the detection stage. Due to the kind of motions present in this sequence (the bottom part of the tower is the first one that appears, as shown in Fig. 6, Frames 3231 and 3628), the template image used by the tracking algorithm was updated every 20 frames in the first 100 frames of the sequence (lines are detected inside the ROI and are used to update the size of the template
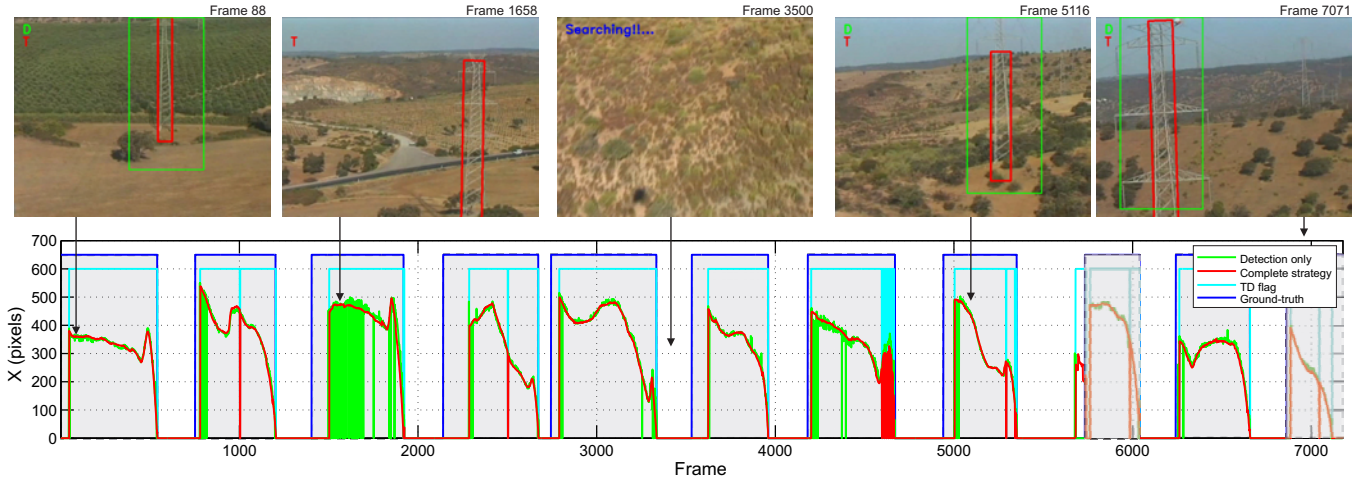
Fig. 7. Comparison of the application of the complete strategy with the application of only the detection stage. The x coordinates of the centers of the ROI found using only the detection stage (green/light line) and using the complete strategy (red/dark line) are plotted. Shadowed areas (below the blue/dark line) represent the frames that contain towers, and the cyan/light line shows the status of the tower detection flag TD (for visualization purpose, TD= 600).

image). This permitted the tower to appear in the FOV of the camera before tracking for large intervals (e.g Fig. 6, Frame 228). After 100 frames, the template image was updated only when the switching criteria defined in Section III-D were not satisfied.

Fig 7 shows a comparison of the application of the complete strategy (tracking+detection, red/dark line) with the application of only the detection stage (green/light line), where the detection stage was applied in each frame of the image sequence. Thumbnail images show some of the results of the comparison (only detection, green/light box; and complete strategy, red/dark box). The green/light line shown in the figure corresponds to the x coordinate of the center of the ROI found by the detection algorithm. The red/dark line represents the x coordinate of the center of the ROI found by the complete strategy (detection+tracking). In this plot, the shadowed areas (below the blue/dark line) represent the frames of the image sequence that contain towers or part of towers (this has been found manually); and the cyan/light line shows the status of the tower detection flag TD (for visualization purposes, the value of this flag was modified to TD= 600).

When only the detection stage was tested, it was observed that, although in some frames (e.g. between 1500 and 2000) the tower detection had some trouble detecting towers (which can be due to the kind of background present in those frames), it detected each of the 11 towers that appeared in the sequence at least once. However, the position of the towers was found in more frames when the complete strategy was used (red/dark line).

Additionally, in Fig. 7 it can be observed that when the complete strategy was tested (red/dark line) the tracking stage estimated the position of the tower most of the time, and the intervention of the detection stage was required only in a few cases, e.g. in Frame 1000 the algorithm switched from the tracking stage (TD= 600) to the detection stage (TD= 0), see cyan/light line, and therefore the tracking

algorithm was reinitialized. The detection and tracking criteria switched between the detection and tracking stages satisfactorily.

In the figure, it can be seen that a false positive found by the detection stage around Frame 5600 makes the tracking algorithm track an area that did not contain any tower. Nonetheless, one of the tracking criteria switched TD= 0, and therefore the detection stage was in charge of reinitializing the tracking task (this was successfully achieved after a few frames).

Finally, Fig. 8 compares the increments in the position of the x and y coordinates of the center of the ROI found when only the detection stage was used (green/light line) with the ones found when the complete strategy was used (red/dark line). In this figure, it can be seen that by integrating the detection and tracking stages, the position of the tower is found smoothly frame-by-frame, which was not the case when the detection stage operated in isolation (the green/light lines show strong frame-to-frame motion).
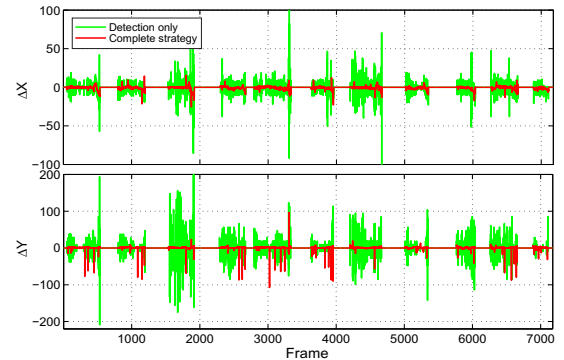

Fig. 8. Comparison of the frame-to-frame motion. The increments in the position of the x and y coordinates of the center of the ROI found when only the detection stage was used (green/light line) are compared with the ones found when the complete strategy was applied (red/dark line). Smoother frame-to-frame motions are found with the complete strategy.

2) Medium Voltage Tower: A second image sequence is used to test the proposed tower detection and tracking strategy. The image sequence contains 4 towers of two different types of medium voltage towers (1639 frames). This sequence represents a more challenging scenario for the tower detection and tracking strategy, not only because of the kind of tower (simple structure), but also because of the kind of motion present in the sequence. Poor image quality, strong camera motions, abrupt changes in the appearance of the tower (the helicopter moves around the tower), strong motion blur, and part of the tower being outside the FOV, are some of the features found in this sequence.

Fig. 9 presents a collection of images illustrating the performance of the proposed tower detection and tracking strategy. In this figure, two types of towers can be seen (e.g. Frame 135 and Frame 1143). Due to the strong motion blur and appearance changes of this sequence, the template image used by the tracking algorithm was updated every 10 frames (lines are detected inside the ROI and are used to update the size of the template image).
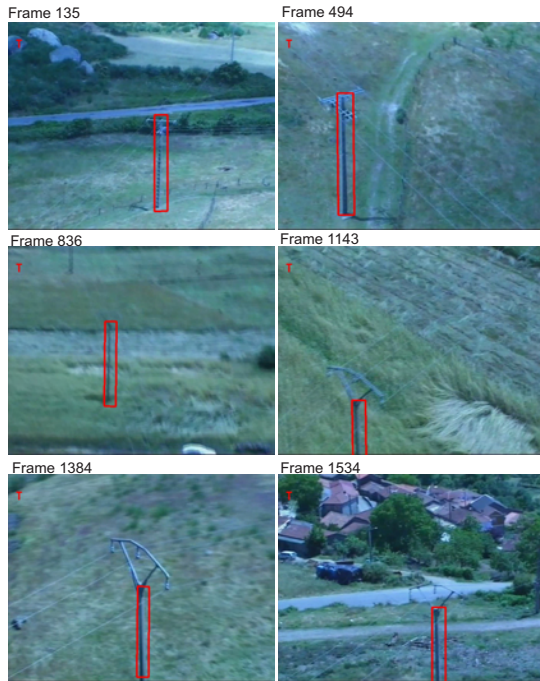

Fig. 9. Sample results of applying the complete detection and tracking strategy in an image sequence with medium voltage towers.

Fig 10 shows the comparison of the application of the complete strategy (red/dark line) with the application of only the detection stage (green/light line). Thumbnail images show some of the results of the comparison (only detection, green/light box; and complete strategy, red/dark box). Comparing both algorithms, it can be seen that when only the detection stage was used, there were several frames in which the tower was not detected (e.g. Frame 305, thumbnail image). However, because of the integration of the tracking and the detection stages, the tower was tracked in almost all the frames of the sequence, overcoming the difficult
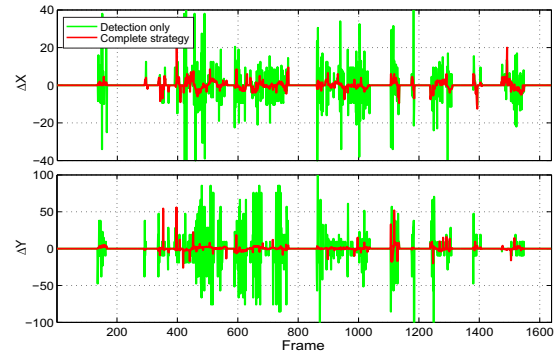

Fig. 11. Comparison of the frame-to-frame motion of the detection and the detection+tracking algorithms for medium voltage tower sequence.

visual conditions of this sequence, such as strong motion blur (e.g. Frames 305 and 805, in the thumbnail images) and changes in appearance (e.g. Frame 145 and Frame 305 correspond to the same tower). Additionally, it is important to highlight the performance of the detection stage in this image sequence, where no false positives were found in images only containing background information.

Finally, Fig. 11 compares the increments in the position of the x and y coordinates of the center of the ROI found when only the detection stage was used (green/light line) with the ones found when the complete strategy was used (red/dark line). The integration of the tower detection and tower tracking stages permit to obtain an estimation of the motion of the tower in the image plane that is smoother than the one obtained when the detection stage was working in isolation.

D. Discussion

Different tests were conducted in order to analyze the performance of the tower detection and tracking strategy that is proposed in this paper for power line inspection with cameras on-board UAVs. The performance of the proposed strategy has been tested in very challenging scenarios that contain different kinds of towers and background, poor image quality, strong motion blur, among other features.

Due to the kind of images used in the test, some assumption had to be done depending on the kind of motions present in the videos. However, it is important to mention that when the system is on-board a UAV with a predefined flight strategy, these assumptions are not required.

During the test, the tower detection stage has shown to be robust when detecting and localizing different kinds of towers in cluttered backgrounds; and have also shown to be robust when classifying background information. The algorithm only estimated a false positive in one frame (the sequence with high voltage towers).

On the other hand, the tower tracking strategy has proved to be robust under adverse conditions, including fast image motion, appearance changes and only partial presence of the towers in the image.

Table III summarizes the results obtained in both sequences. In the high voltage sequence (HV), of 4993 frames
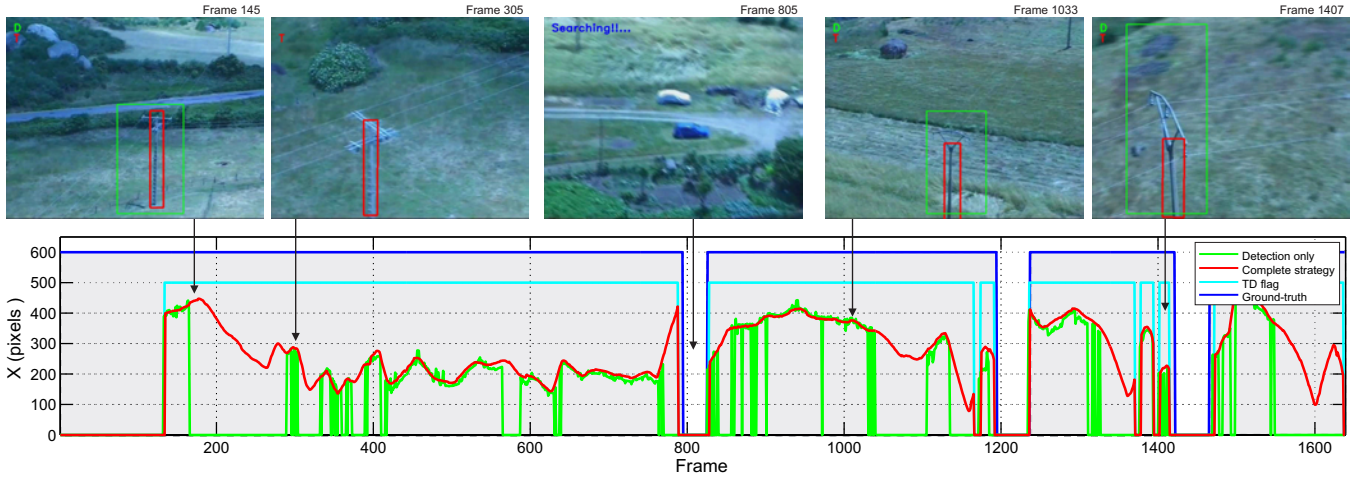
Fig. 10. Comparison of the application of the complete strategy with the application of only the detection stage in the image sequence with medium voltage towers.

that contain towers (this is an approximated value estimated analyzing the image sequences), the complete strategy (detection+tracking) detected towers in 4323 frames. Conversely, when only the detection stage was used, the position of the towers was found in 4148 frames. On the other hand, in the medium voltage sequence (MV), of 1522 frames that contain towers, the tower detection and tracking strategy detected towers in 1334 frames, whereas the detection stage detected towers in 794 frames. From this table, it can be seen that in both tested sequences the complete strategy has a better performance when estimating the position of the tower than when only the detection stage was used.

TABLE III

TRACKING STAGE AND DETECTION STAGE COMPARISON

Image Sequence	# Frames	Tower present	Complete strategy	Detection only
HV	7177	4993	4323	4148
MV	1639	1522	1334	794

As can be seen Table III, neither the detection stage nor the complete strategy detected the tower in all the frames that contained towers. This is mainly because in some of the frames the position and size of the tower do not satisfy the detection and tracking criteria used to monitor the performance of the algorithms. Therefore, in those frames the algorithm considers that there is not tower. This is mainly because the information present in those frames is not relevant for visual inspection purposes.

Fig. 12 shows some of the images with some false positive detection and some tracking errors. Frames 826, 1184, 4663, and 5681 correspond to some false positive detections (green/light box). As can be seen in some of those images, the tracking algorithm estimated the correct position of the tower (red/dark box). Most of the false positive detections occurred in the medium voltage sequence. This is mainly because of the structure of the towers and the poor quality of the images, that make it difficult to identify a tower (an example of this can be seen in Frame 1191).
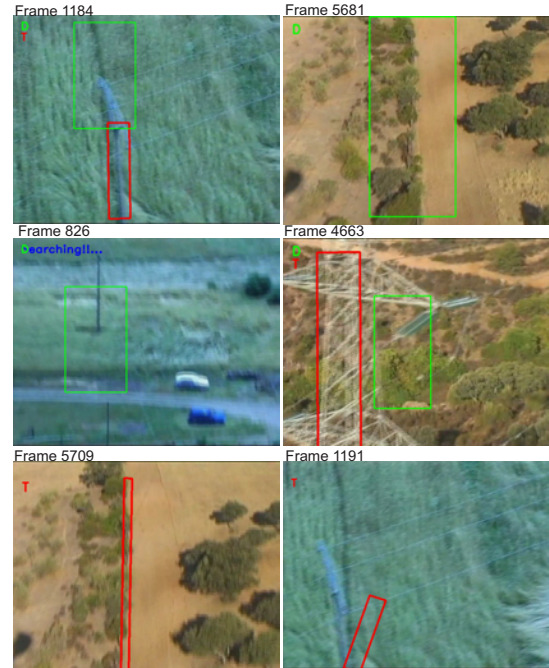

Fig. 12. False positive detections and tracking errors.

On the other hand, Frames 1191 and 5709 show two examples of the few errors obtained when the complete strategy was tested. In Frame 5709, the wrong ROI detected by the detection stage in Frame 5681 made the tracking algorithm track a ROI that does not contain a tower. Additionally, because the ROI contains a line, it took some frames to the tracking algorithm to trigger the detection stage. Conversely, a possible cause of the error of Frame 1191 is the poor quality of the image and strong motion blur, which made the tracking algorithm estimate the position of the tower wrongly.

Nevertheless, it is important to notice that in spite of the errors shown in Fig. 12, in general the proposed tower detection and tracking strategy has been shown to be of a

standard that is appropriate to track electric towers under very challenging circumstances.

V. CONCLUSIONS AND FUTURE WORK

Electric towers are a key component of the power line infrastructure, all the more so because they consist of a large set of items extremely critical for safe and consistent power supply. This paper presents a strategy for detecting and tracking the electric towers during non-intensive aerial inspections. An approach combining machine learning methods, for autonomous tower detection, and computer vision techniques, for subsequent tracking of towers, was presented.

A two-class MLP (multilayer perceptron) was trained for Tower-Background classification and was applied as a sliding window detector. In an independent evaluation of the detector, highly encouraging results were obtained. The application of a pre-processing step showed a marked improvement in the performance of the tower detection stage.

The output of the detector triggers the tracker. The tower tracking strategy is based on a hierarchical tracking-by-registration strategy, which uses the Hierarchical Multi-Parametric and Multi-Resolution Inverse Compositional Algorithm [1].

The evaluation results show that the integration of the detection and tracking stages, the proposal of this paper, led to a very robust strategy for detecting, localizing and tracking the towers in very challenging conditions.[1]

The proposed strategy is of special significance for UAV-based inspections. Successful detection and tracking of the towers can be used for steering an on-board camera to keep and maintain the tower in the FOV, and a confident localization of the tower can be used for "closing the navigation loop". Therefore, an immediate continuation of the presented work will be to apply and test the proposed approach on a UAV platform.

ACKNOWLEDGEMENTS

REFERENCES

[1] C. Martínez, I. F. Mondragón, P. Campoy, J. L. Sánchez-López, and M. A. Olivares-Méndez, "A hierarchical tracking strategy for vision-based applications on-board uavs," *Journal of Intelligent & Robotic Systems*, pp. 1–23, 2013.

[2] R. Aggarwal, A. Johns, J. Jayasinghe, and W. Su, "An overview of the condition monitoring of overhead lines," *Electric Power systems research*, vol. 53, no. 1, pp. 15–22, 2000.

[3] D. Jones and G. Earp, "Camera sightline pointing requirements for aerial inspection of overhead power lines," *Electric Power Systems Research*, vol. 57, no. 2, pp. 73–82, 2001.

[4] J. Katrasnik, F. Pernus, and B. Likar, "A survey of mobile robots for distribution power line inspection," *Power Delivery, IEEE Transactions on*, vol. 25, no. 1, pp. 485–493, 2010.

[5] A. Pagnano, M. Höpf, and R. Teti, "A roadmap for automated power line inspection. maintenance and repair," *Procedia CIRP*, vol. 12, pp. 234–239, 2013.

[6] D. Jones, "Power line inspection-a uav concept," in *Autonomous Systems, 2005. The IEE Forum on (Ref. No. 2005/11271)*. IET, 2005, pp. 8–pp.

[7] P. Campoy, P. J. Garcia, A. Barrientos, J. del Cerro, I. Aguirre, A. Roa, R. Garcia, and J. M. Muñoz, "An stereoscopic vision system guiding an autonomous helicopter for overhead power cable inspection," in *Robot Vision*. Springer, 2001, pp. 115–124.

[8] S. Montambault, J. Beaudry, K. Toussaint, and N. Pouliot, "On the application of vtol uavs to the inspection of power utility assets," in *Applied Robotics for the Power Industry (CARPI), 2010 1st International Conference on*. IEEE, 2010, pp. 1–7.

[9] Z. Li, Y. Liu, R. Walker, R. Hayward, and J. Zhang, "Towards automatic power line detection for a uav surveillance system using pulse coupled neural filter and an improved hough transform," *Machine Vision and Applications*, vol. 21, no. 5, pp. 677–686, 2010.

[10] J. Sawada, K. Kusumoto, Y. Maikawa, T. Munakata, and Y. Ishikawa, "A mobile robot for inspection of power transmission lines," *Power Delivery, IEEE Transactions on*, vol. 6, no. 1, pp. 309–315, 1991.

[11] K. Toussaint, N. Pouliot, and S. Montambault, "Transmission line maintenance robots capable of crossing obstacles: State-of-the-art review and challenges ahead," *Journal of Field Robotics*, vol. 26, no. 5, pp. 477–499, 2009.

[12] J. Zhao, R. Guo, L. Cao, and F. Zhang, "Improvement of linerover: A mobile robot for de-icing of transmission lines," in *Applied Robotics for the Power Industry (CARPI), 2010 1st International Conference on*. IEEE, 2010, pp. 1–4.

[13] J. Katrasnik, F. Pernus, and B. Likar, "New robot for power line inspection," in *Robotics, Automation and Mechatronics, 2008 IEEE Conference on*. IEEE, 2008, pp. 1195–1200.

[14] I. Golightly and D. Jones, "Corner detection and matching for visual tracking during power line inspection," *Image and Vision Computing*, vol. 21, no. 9, pp. 827–840, 2003.

[15] C. Whitworth, A. Duller, D. Jones, and G. Earp, "Aerial video inspection of overhead power lines," *Power Engineering Journal*, vol. 15, no. 1, pp. 25–32, 2001.

[16] C. Sun, R. Jones, H. Talbot, X. Wu, K. Cheong, R. Beare, M. Buckley, and M. Berman, "Measuring the distance of vegetation from powerlines using stereo vision," *ISPRS journal of photogrammetry and remote sensing*, vol. 60, no. 4, pp. 269–283, 2006.

[17] W. Cheng and Z. Song, "Power pole detection based on graph cut," in *Image and Signal Processing, 2008. CISP'08. Congress on*, vol. 3. IEEE, 2008, pp. 720–724.

[18] J. Tilawat, N. Theera-Umpon, and S. Auephanwiriyakul, "Automatic detection of electricity pylons in aerial video sequences," in *Electronics and Information Engineering (ICEIE), 2010 International Conference On*, vol. 1. IEEE, 2010, pp. 342–346.

[19] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 1. IEEE, 2005, pp. 886–893.

[20] V. Ferrari, L. Fevrier, F. Jurie, and C. Schmid, "Groups of adjacent contour segments for object detection," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 30, no. 1, pp. 36–51, 2008.

[21] H. Harzallah, F. Jurie, and C. Schmid, "Combining efficient object localization and image classification," in *Computer Vision, 2009 IEEE 12th International Conference on*. IEEE, 2009, pp. 237–244.

[22] M. Irani and P. Anandan, "About direct methods," in *Vision Algorithms: Theory and Practice*, ser. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, 2000, vol. 1883, pp. 267–277.

[23] S. Baker and I. Matthews, "Equivalence and efficiency of image alignment algorithms," *Proceedings of the 2001 IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 1090 – 1097, December 2001.

[24] C. H. Anderson, J. R. Bergen, P. J. Burt, and J. M. Ogden, "Pyramid methods in image processing," RCA Engineer, vol. 29, no. 6, pp. 33-41, Nov./Dec., 1984.

[25] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. New York, NY, USA: Cambridge University Press, 2003.

---

[1]A video demonstration of the reported results has been made available at: http://youtu.be/uE4TQ7CocJM