A STEREO APPROACH TO WILDFIRE SMOKE DETECTION: THE IMPROVEMENT OF THE EXISTING METHODS BY ADDING A NEW DIMENSION

Toni Jakovčević, Marin Bugarić, Darko Stipaničev

Faculty of Electrical Engineering, Mechanical Engineering and Naval Architecture
University of Split, R. Boškovića 32, 21000 Split, Croatia
e-mail: {toni.jakovcevic, marin.bugaric, dstip}@fesb.hr

Abstract. In this paper, we present a novel approach to visual smoke detection based on stereo vision. General smoke detection is usually performed by analyzing the images from remote cameras using various computer vision techniques. The literature on smoke detection shows a variety of approaches, and the focus of this paper is the improvement of the general smoke detection process by introducing stereo vision. Two cameras are used to estimate the distance and size of the detected phenomena based on stereo triangulation. Using this information, the minimum size and overall dynamics of the detected regions are further examined to ensure the elimination of false alarms induced by various phenomena (such as the movement of objects located at short distances from the camera). Such false alarms could easily be detected by the proposed stereo system, allowing the increase of the sensitivity and overall performance of the detection. We analyzed the requirements of such system in terms of precision and robustness to possible error sources, especially when dealing with detection of smoke at various distances from the camera. For evaluation, three existing smoke detection methods were tested and the results were compared to their newly implemented stereo versions. The results demonstrated better overall performance, especially a decrease in false alarm rates for all tested methods.

Keywords: Smoke detection, artificial intelligence, stereo vision, false alarms, wildfires

Mathematics Subject Classification 2010: 68T45, 68U10

1 INTRODUCTION

Wildfires, unlike many other natural disasters, are perhaps the only ones that can be largely prevented. Nevertheless, the consequences of uncontrolled wildfires are catastrophic and can lead to significant material damage and can have detrimental impacts on both human safety and health. Regardless of whether the wildfires started accidentally or caused by arson, the hazard can be avoided if they are promptly identified and quickly extinguished.

A lot of efforts have been invested in the early detection of wildfires. Traditional detection is usually based on observers who monitor the surrounding environment in search for a smoke. This is due to the nature of wildfires, where in most cases smoke is visible long before the flame. This is particularly evident for environments with dense vegetation such as forest areas, where fire is not visible until it catches the crowns. The observers are located on observation towers positioned at carefully chosen locations, such as hills, in order to have a better view of the surroundings. Unfortunately, wildfires often occur during extreme conditions such as drought and high temperatures, leading to concentration difficulties and a decreased ability to focus on the actual recognition of smoke.

This issue was partly solved by introducing a camera-based surveillance, allowing the observer to control several observation posts from a single remote location. Unfortunately, practice has shown that these systems still require a long-term attention. To further improve the effectiveness of such systems, automatic smoke detection methods have been developed. Reliability of these methods relies on advanced computer vision algorithms that take into account many different smoke characteristics in order to properly identify the smoke in the image taken from the camera. The system raises an alarm if it determines the presence of smoke in the image. However, a final confirmation by a human operator is still required to distinguish real threats from false alarms.

There is a relation between false alarm and correct detection rates for a given smoke detection system. Although most existing systems have the ability to change the detection sensitivity using detection parameters, this relation between false alarms and correct detections is not significantly affected. An increase in detection sensitivity usually leads to a higher number of correct detections, but unfortunately also to a higher number of false alarms. On the other hand, lowering the detection sensitivity can result in a missed detection, making the system unusable and unreliable. It would be of great benefit to find a solution to reduce the number of false alarms while simultaneously maintaining the acceptable number of correct detections.

In this paper we propose the improvement of the existing smoke detection methods by introducing stereo vision to the detection system. Application of stereo vision in smoke detection systems has already proven to be useful, as shown in [1, 2]. However, in those solutions, stereo vision is primarily used to extract foreground objects from the background. Foreground objects are then further analyzed: i.e., in [1] wavelet transform and discrete cosine for feature extraction and recognition based

on fuzzy-neural networks is used, while in [2] additional image features are extracted (contrast, brightness, edge strength, etc.) in order to validate smoke regions. This is different from our approach, where we use stereo vision to estimate the distance to the object or phenomenon visible in the image, what is further used to analyze the detected regions; i.e., to check the minimum size and overall dynamics (changes in size over time) of the detected candidate region. If those regions do not meet specified requirements, they can be discarded as false alarms. Please note that except from introducing this feature to the detection process, the original smoke detection methods are not modified in any way. However, lower false alarms rate achieved by the proposed improvement allows us to raise the detection sensitivity. In other words, it is possible to achieve not only lower rate of false alarms, but also to increase the rate of correct detections as well as the coverage area of the used smoke detection method.

Depth information of the scene visible in the image has also been used to improve existing smoke detection algorithms in recent work presented by Bugarić et al. [3]. However, in this solution distances are estimated using the precise virtual terrain model. The main disadvantage of this system is that previous calculations are mandatory to operate correctly. Calculating a depth map is done for the entire image, which is time-consuming and can last up to ten minutes for high-resolution images. Therefore, such calculations are done only for predefined camera preset positions.

The approach presented in this paper uses stereo vision to process only the candidate regions detected as smoke. In this way we eliminate the need for calculating the depth map for the entire image, meaning that we do not significantly influence the execution time of the original smoke detection method. Also, with this approach, there is no need for using predefined preset positions and the camera can move freely. Please note that using stereo vision, the calculated distance represents the actual distance from the camera to the detected phenomenon, rather than the distance to the terrain behind it, as it is the case in [3].

The rest of the paper is organized as follows: in Section 2 we give an overview of existing visual smoke detection systems where we investigated the most common phases for the detection process. In Section 3 we present our solution to visual smoke detection based on stereo vision, Section 4 deals with possible improvements over standard smoke detection approaches. Finally, a thorough evaluation is carried out in Section 5 where we implemented stereo versions of three existing visual smoke detection methods and compared them to their standard versions.

2 OVERVIEW OF VISUAL SMOKE DETECTION

There are various approaches to visual smoke detection, and in this section we will cover the most common methods and aspects of the detection process. The research in this field has begun over twenty years ago, with one of the first papers [4] dealing with smoke detection based on a live stream from a surveillance camera. The de-

veloped systems have been constantly improving since that period, and the field is very alive with new methods constantly emerging. However, most of the methods share several phases that are common to the general smoke detection process. We can divide the detection process in several common phases and cover each phase separately. It is important to emphasize that these phases do not have to be executed sequentially, or in a specific order. In some approaches they are executed simultaneously, and this process is method dependent. Most commonly used phases in smoke detection are: motion detection, region analysis, dynamics analysis and the decision phase. In the following subsections we will briefly cover these phases.

2.1 Motion Detection

Motion detection is the most common phase in smoke detection. Wildfire smoke is dynamic, and this property is used to isolate only the moving regions from the entire scene. This phase generally acts as filter for subsequent phases to reduce the amount of data for further computation, so only the detected regions are forwarded for further analysis. There are many approaches to motion detection for smoke detection purposes such as adaptive background estimation [5, 6], block mean difference [7, 8] or motion history image [9, 10]. Adaptive background estimation approaches in [5, 6] are similar and use a background estimation model based on rules for stationary and moving pixels. In the case that the pixel intensities significantly deviate from the background model they are considered as moving pixels and forwarded to the subsequent phases for further analysis. Another motion detection approach is described in [11] where the background model is initialized dividing the input in 16×16 size blocks. After the setup phase, the model is updated using a selective temporal median with a fixed k-sized circular buffer. The difference between the input image and the background model is computed and then binarized using a low and high thresholds to identify small and high intensity variations. Those pixels that are present in both of the masks generated by low and high thresholds are considered as moving pixels. Extracted objects are then validated jointly using color shape and gradient information to remove noise and artifacts. Further processing phases involve shadow and ghost removal.

Motion detection approach using motion history image described in [9] represents motion in successively layered image differences. Moving objects create silhouettes that represent patterns of motions. The intensity in a motion history image (MHI) represents the recency of motion in the observed scene. In this way, the motion from several frames can be encoded in a single image. Using this approach, it is possible to capture the gradient of smoke motion including orientation and direction.

2.2 Region Analysis

Another very important phase in the general smoke detection process is the region analysis. This phase is often executed after motion detection, where region candidates are extracted. Regions are analyzed based on different smoke characteristics

such as color, texture, shape and size. Different methods use different approaches, but it is generally a combination of several characteristics. However, all of the methods rely on color as one of the most important features. When dealing with color analysis, one of the first steps is to choose a color space for analysis. Research described in [12] deals with the effect of various color spaces on the performance of different classifiers in smoke detection. One of the goals of the research was to find a particular color space with highest separability between smoke and non-smoke pixels. Different color spaces were used in the analysis, such as RGB, YCrCb, CIELab, HSI, and a HS'I which is an derivative of HSI. The results suggest that the performance is a distinctive feature of the classifier itself, rather than the classifier-color space combination, and that the favorable color space candidates are HSI and its derivative, as well as RGB color space. Regardless of the color space, all of the methods take into account specific chromatic characteristics of smoke. Smoke color varies based on fuel type and moisture content but it is generally manifested as a light to dark shade of gray. This means that the smoke pixels are positioned diagonally in the RGB color space and it is possible to dismiss the pixels that have high deviation from the diagonal of the color space as implemented in [8, 13]. Another smoke-specific feature is low chrominance in the smoke affected region. Detection of a significant drop in chrominance could be a possible indicator of smoke appearance. Another step in region analysis is using information about the texture of a given region. There are various approaches to texture analysis, such as wavelet analysis [14, 15] or the gray-level co-occurrence matrix (GLCM) [16]. Another important factor in region analysis is the information about the morphological characteristics of the region. Smoke regions have a rather convex contour based on irregular shape with erratic silhouette [5]. Most common approach to isolate these types of regions is to calculate the disorder parameter and compare it to the reference values [17].

2.3 Dynamics Analysis

Another common phase in smoke detection process is the analysis of the dynamics of the candidate regions. The candidate regions are tracked over a certain period of time to ensure they exhibit smoke-like behavior. Regions that do not conform to motion characteristics of smoke can be rejected in this phase of the process. One of the basic characteristics is the growth rate of smoke. Different methods use certain thresholds to isolate only the regions that satisfy a set of predefined conditions [18, 6]. Another aspect is the direction of smoke movement. Smoke usually exhibits upward as well as lateral motion, however, this factor heavily impacted by the wind speed and direction and is rather difficult to predict. Work presented in [19] deals with the analysis of the direction of smoke motion using accumulative motion orientation. The aim of the analysis is to discover regions that are moving upwards due to high temperatures. The method involves calculation of the temporal motion orientation histogram. By removing the regions that do not exhibit growth and gradual upward motion it is possible to reduce the overall number of false alarms and increase the performance of the detection method.

2.4 Decision Phase

The decision phase is the final phase of the detection process. All the information accumulated through the previous phases of detection is now taken into account in order to make the final decision about raising an alarm. There are various approaches to the decision-making based on the data from the previous phases of detection, such as the Bayesian approach [11], neural networks [6, 20], random forests [21], support vector machine [22], the mechanism of thought [23], and others. Based on the available information and the specific decision process, the system makes the final decision whether to raise an alarm for the given situation.

3 INTRODUCING STEREO VISION TO SMOKE DETECTION

Smoke is a phenomenon that has no clearly defined characteristics such as shape, color, etc. This is precisely the reason why there are many different approaches to smoke detection. However, analyzing the size of the smoke region candidates rarely occurs among these approaches. This is due to limitations of the systems that use a single camera, where the only option is to express this size in the number of pixels. Using only this information, it is difficult to estimate the actual size of the smoke phenomenon in the real world, as it largely depends on the camera parameters and the distance of the phenomenon from the camera.

Stereo vision triangulation allows us to estimate a three-dimensional position of the detected smoke in the real world, and also to estimate the real-world size of that region and express it in standard units of measurement. In this way we can further analyze the candidate regions detected as smoke based on their real-world sizes, and thus reduce the number of false alarms. One such example would be identifying false alarms caused by the uncontrolled movement of the vegetation in the close vicinity of the camera based on the estimated size and overall dynamics of the detected objects. Moreover, eliminating such types of false alarms allows us to increase the detection sensitivity, eventually leading to the increase in correct detections rate and coverage area of the used smoke detection method.

Any existing smoke detection system could be upgraded with this stereo vision approach. Given that the smoke detection requires an estimation of relatively large distances, a stereo vision system with a wide baseline is required. In the following sections we analyze the problems that may arise and describe the guidelines that have to be followed when upgrading the existing smoke detection systems.

3.1 Wide-Baseline Stereo Vision

In order to improve smoke detection performance, a stereo system should correctly match the images captured by the cameras and obtain accurately triangulated three-dimensional data. It is important to emphasize that the occurrence of smoke is usually located several hundred meters away from the observation post where the

system is installed. Stereo systems capable of estimating such relatively large distances have already been proposed [24, 25], however a study concerning advantages and disadvantages of such a system as a part of a smoke detection was not carried out to date.

Our system requires two cameras with a known horizontal displacement, whose optical axes are parallel. In order to maintain the accuracy of a stereo system, the horizontal displacement between these two cameras should be relatively wide. The actual smoke detection is conducted on the images taken by one of the stereo cameras (either left or right). Stereo vision techniques are applied only for the candidate regions detected as smoke as the additional verification phase. In our solution, the process of determining the distance of the detected phenomenon from the camera is divided into two steps:

- Selecting a point that represents a smoke region candidate in the original image and finding a corresponding point in the image of the other camera (correspondence problem), and
- three-dimensional reconstruction (stereo triangulation).

Each of these steps is further explained in the following subsections.

3.1.1 Correspondence Problem

The correspondence problem is the problem to match image points from two stereo images which are projections of the same point in a three-dimensional space. The main requirement while taking a pair of stereo images in our solution is that the pictures are taken at the same time instant. Smoke is a dynamic phenomenon, constantly changing in shape and size. If this requirement is not met, projection of the smoke on the image plane can be significantly different in two stereo images, making the matching process difficult. As we show later in this section, even a single pixel error in this phase could lead to significant errors when estimating the real-world sizes of the smoke phenomenon. Please also note that any difference in perspective or lighting of two stereo cameras, or any object occlusions can further accentuate the correspondence problem.

Stereo cameras should, therefore, be mutually synchronized in order to capture the images at the same time instant [26]. Automatic brightness adjustment should also be avoided, as it may result in different lightning in two stereo images.

The points of interest would be the points that the smoke detection algorithm determines as the potential candidates. The detection algorithm isolates the regions of interest, and for each region a single point is selected for correspondence. The selected point basically represents the geometric center of the region. Now, the task is to find the corresponding point in the other image. There is a significant number of existing algorithms for stereo correspondence. However, the main problem is not to generate the disparity map for the given scene, but rather to calculate the actual depth for a very sparse set of points that represents the candidate regions in each frame.

One possible way of achieving this task is to use normalized cross correlation. It is a rather straightforward approach where a region around the given point is used as a template. The aim is to find the best matching position of that template in the second image. The search is not performed on the complete second image, but rather on a limited region that covers the space were the corresponding point could be in the second image given the constraints of the geometry of the stereo system. Given the possible variation in brightness due to exposure or lightning conditions it is necessary to perform the normalization of the regions. The point with the highest correlation coefficient is used as the corresponding point in the second image. The process of normalized cross correlation could be computationally demanding when performed on large sets of correspondence points, but for a very sparse set that is generally obtained from the candidate regions the computing load is negligible. We can examine this in more detail. Specific implementation of normalized correlation calculation is based on [27] as follows

$$\gamma(u,v) = \frac{\sum_{x,y} \left[f(x,y) - \bar{f}_{u,v} \right] \left[t(x-u, y-v) - \bar{t} \right]}{\left\{ \sum_{x,y} \left[f(x,y) - \bar{f}_{u,v} \right]^2 \sum_{x,y} \left[t(x-u, y-v) - \bar{t} \right]^2 \right\}^{0.5}} \tag{1}$$

where f is the second image where we search for a match, $\bar{t}$ is the mean of the template, and $\bar{f}_{u,v}$ is the mean of $f(x,y)$ in the region under the template. This calculates the normalized correlation coefficient for the template placed at a point (u,v) in the second image. We can see that the time complexity of calculation for a single point would be $O(M_t N_t)$ where M_t and N_t represent the size of the template. In case we would like to calculate the correspondence of each point in the image of size $M_I \times N_I$ the time complexity would be $O(N_I M_I M_t N_t)$. It is evident that the computational cost of type of calculation is rather high, so it is necessary to perform the calculation only for a set containing the points that represent the candidate regions in order for the algorithm to work in real-time. For example, we can take that the time constraint for real-time processing is 1 frame per second in order to capture smoke dynamics. Average time to calculate normalized cross correlation for a single point, where the template dimensions are 40×20 pixels, and the search space in the second image is 900×100 pixels, on a machine with $3\,\mathrm{GHz}$ CPU is approximately $14\,\mathrm{ms}$. This implies that in order for the algorithm to be able to work in real time we must use a sparse set of correspondence points that are most probable candidates for detection.

3.1.2 Stereo Triangulation

The distance of the observed object or phenomenon from the camera is computed using the disparity of two corresponding pixels. In our model, each camera is reasonably approximated by a pinhole camera model, thus ignoring camera lens distortion and other optical nonlinearities. We consider lens to be a point through which all incoming rays of light pass. This means that all objects, regardless of their distance from the camera, project to a single point on the image plane.
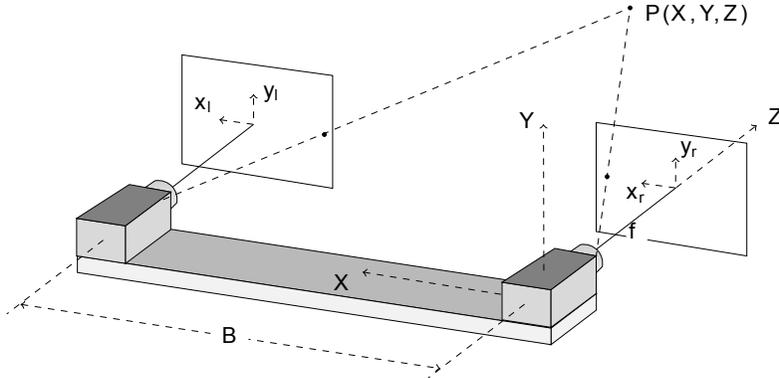
Figure 1. Stereo system

Figure 1 illustrates the configuration of a stereo system that we use to improve the existing smoke detection methods. Two cameras that have parallel image planes are separated in the X direction, and the baseline length is denoted by B. Both cameras share the same focal length that we denote by f. The focal point of the right camera is chosen as the origin of the three-dimensional world coordinate system. Please note that, in order to minimize the measurement inaccuracy, these two cameras have to be carefully aligned and well calibrated. Also, the cameras should be securely mounted to ensure that the alignment is persevered with continuous use.

In this model, a randomly chosen point $P(X, Y, Z)$ in a three-dimensional world is projected onto two camera image planes. These projections in the left and right image planes are represented by (x_l, y_l) and (x_r, y_r), respectively.

The relation of the point $P(X, Y, Z)$ with respect to its image projections is given by:

$$x_r = \frac{fX}{Z}, \quad x_l = \frac{f(X - B)}{Z}. \tag{2}$$

Therefore, using simple trigonometry, we can estimate the depth Z of the point $P(X, Y, Z)$:

$$Z = \frac{fB}{x_r - x_l}. \tag{3}$$

From Equation (3) we see how the distance from the camera can be computed using the disparity $x_r - x_l$. It can be seen that y_l and y_r coordinates do not influence the actual calculation of the depth information. Please note that the aforementioned model does not take into account the following irregularities: different focal lengths of left and right camera, differences in principal point coordinates, skew and distortion parameters. Therefore, in order to use Equation (3) in practice, inputs into this equation should first be normalized. However, even without such irregularities, Equation (3) will not always provide us with the correct depth information. This occurs for several reasons, so in the following of this chapter we will investigate pos-

sible errors and discuss how they affect the proposed improvement of the existing smoke detection methods.

3.2 Different Types of Errors

First, we must accept that the depth resolution of a stereo vision system is limited due to the discrete nature of the imaging system [28, 24]. The projection of the point $P(X, Y, Z)$ onto the image planes is approximated to the nearest pixels with coordinates $(\hat{x}_l, \hat{y}_l)$ and $(\hat{x}_r, \hat{y}_r)$. The error that is a result of this approximation is referred to as a discretization error.

The discretization error generates an uncertainty polyhedron when we perform the stereo triangulation. Due to the aforementioned approximation, the actual distance cannot be accurately determined, since it lies somewhere inside this polyhedron. Figure 2 illustrates the discretization error in two dimensions, where uncertainty areas are represented as diamond shapes.

The discretized image points that cause the error are at most within half a pixel of the actual projection, that is:

$$\hat{x}_r = x_r \pm \frac{x_{pix}}{2}, \quad \hat{x}_l = x_l \pm \frac{x_{pix}}{2} \tag{4}$$

where x_{pix} represents the distance between two adjacent pixels along the X direction. From this it follows that the maximum value of the observed depth $\hat{Z}$ can be calculated using Equation (5):

$$\hat{Z} = \frac{Z}{1 \pm Z\frac{x_{pix}}{fB}}. \tag{5}$$

It can also be shown that the maximum value of the discretization error (Z_{disc_err}) is given by:

$$Z_{disc\_err} = |Z - \hat{Z}| = Z^2 \cdot \frac{x_{pix}}{fB \pm Zx_{pix}}. \tag{6}$$

From Equation (6) we can see that the discretization error (Z_{disc_err}) increases as the depth Z increases. Several other interesting facts useful for a better estimation of relatively large distances arise from this equation as well. First, since the focal length of the calibrated camera is not a parameter that can be easily modified, from Equation (6) it follows that we can improve distance estimation by increasing the length of the baseline B. Second, image resolution affects the discretization error as well, since it defines the distance between two adjacent pixels (x_{pix}). From Equation (6) it is obvious that the higher the resolution of the image, the smaller the discretization error.

Nevertheless, it is important to notice the expression $\pm Zx_{pix}$ in the denominator of Equation (6). Uncertainty areas represented as diamond shapes in Figure 2 are bounded by rays that go through the middle of each pixel in the image plane. Therefore, the distance from the intersection of rays to the center of the diamond

shape corresponds to the maximum discretization error. Recall, the maximum discretization error is achieved by moving half a pixel in a certain direction (in both left and right image planes). As seen from Figure 2 this movement can result in the estimated distance that is larger or smaller then the actual distance. Therefore, the discretization error is measured from the intersection of rays to the center of the diamond shape that is either further away or closer to the stereo cameras. Let us denote these errors as a maximum positive and a maximum negative discretization error, respectively. Maximum positive discretization error is defined by the negative prefix in the expression $\pm Z x_{pix}$, while the maximum negative discretization error is defined by the positive prefix. Maximum discretization error (either positive or negative) represents a worst case scenario, and in most cases the value of the discretization error is smaller. It can also be noticed that the maximum positive discretization error is larger than the maximum negative discretization error.
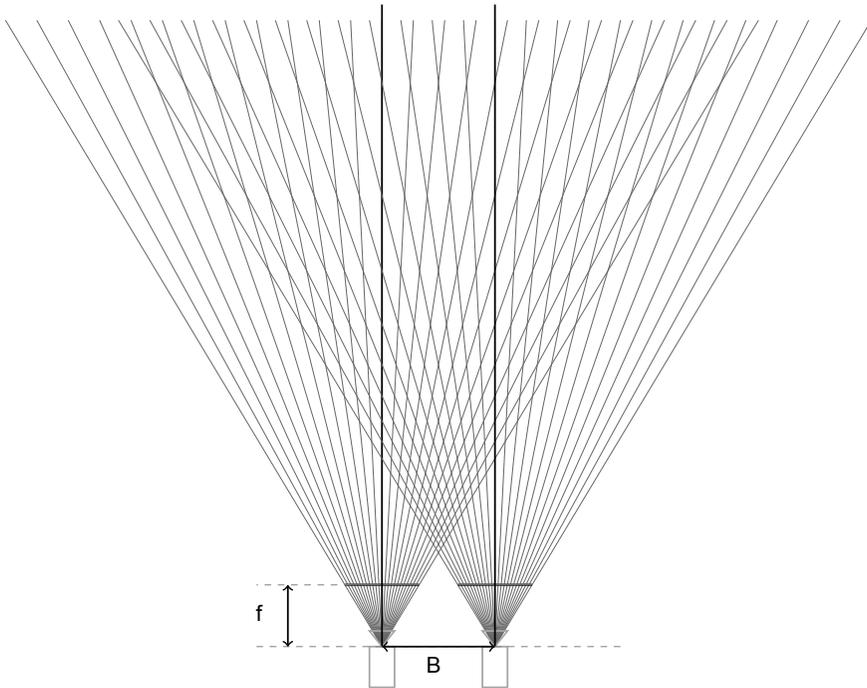

Figure 2. The discretization error as the result of the discrete nature of the imaging system. Diamond shapes represent uncertainty areas, i.e., the system returns the same distance for all the points inside the same diamond region.

It is also very important to consider possible camera alignment errors and their impact on the depth estimation. Once the stereo rig is constructed, the cameras are calibrated in order to obtain intrinsic and extrinsic camera parameters. Using the calibration information we can model the possible misalignments of the individual

cameras relative to each other. However, the perfect calibration is not possible and a certain degree of error will still be present. Additionally, the camera setting is prone to slight misalignment due to continued usage and transport, so we also have to take into account the possible errors that arise from this sort of misalignment. We will consider four types of errors; e.g. error due to roll between cameras, error due to pitch between cameras, error due to yaw between cameras and finally the error due to lens distortion. These types of errors are explained in detail in [29] so we will only cover them briefly. The first type of error is the depth error due to roll of the second camera, while the first camera is correctly aligned. Using trigonometry relations we can arrive at the the following equation:

$$Z_{roll\_err} \simeq Z\frac{X_2(cos\theta - 1)}{B} \tag{7}$$

where Z_{roll_err} is the depth error, Z is the true depth of the object, X_2 is the true 3-D coordinate of the object in the x dimension in the coordinate frame of the second camera, θ is the roll angle of the camera, and B is the baseline.

Another type of possible error arises when the second camera is not on the same level as the first, and rotates about a line which is parallel to the bar (pitch error). This type of error can be calculated as follows:

$$Z_{pitch\_err} \simeq -\frac{1}{2}\frac{X_2 Z(\tan\alpha)^2}{B} \tag{8}$$

where α is the pitch angle between the two cameras.

The third type of error arises when the second camera is rotated about an axis perpendicular to the epipolar plane and through the center of projection (yaw error). This type of error can be approximated using the following equation:

$$Z_{yaw\_err} \simeq -\frac{\tan\beta(Z^2 + X_2^2)}{B} \tag{9}$$

where β is the yaw angle between the two cameras. It is important to emphasize that this type of error is the most common, and contributes the most to the error in measured depth.

Another type of error arises from the possible lens imperfections. Although we use a pinhole camera model, we can also consider the impact of this type of error in real world applications. As covered in [29], using the radial lens distortion model, the depth error can be calculated as

$$Z_{lens\_err} = Z - \frac{1}{\frac{1}{Z} - \frac{f^2}{Z^3 B}[\kappa_2 X_2^3 - \kappa_1(X_2 - B)^3]} \tag{10}$$

where κ_1 and κ_2 are the lens distortion coefficients.

Now, the accumulated error can be approximated as:

$$Z_{acc\_err}(Z, \alpha, \beta, \theta, \kappa) \simeq Z_{disc\_err}(Z) + Z_{roll\_err}(\theta)$$

$$+ Z_{pitch\_err}(\alpha) + Z_{yaw\_err}(\beta) + Z_{lens\_err}(\kappa). \qquad (11)$$

We can compare the effects of different types of errors on the accumulated error. Figure 3 a) shows the comparison of the absolute yaw error and the discretization error for variable distance of the observed object Z and yaw angle β.
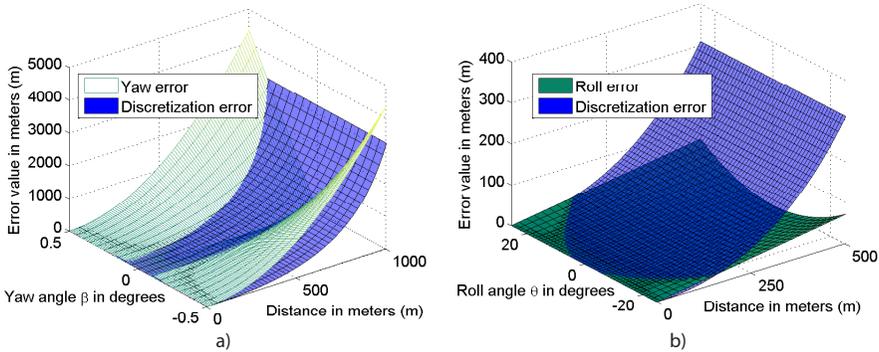

Figure 3. a) Absolute yaw error $|Z_{yaw\_err}|$ (green), and maximum absolute discretization error Z_{disc_err} (purple), b) absolute roll error $|Z_{roll\_err}|$ (green), and maximum absolute discretization error Z_{disc_err} (purple) (for a stereo system with $B = 2\,\mathrm{m}$, $f = 4.655\,\mathrm{mm}$, $x_{pix} = 0.0071\,\mathrm{mm}$, and $X2 = 3\,\mathrm{m}$)

When calculating the discretization error we use the maximum positive discretization error (the expression $\pm Zex$ in Equation (6) with a negative prefix) since it generates a greater overall error. It is possible to see from the figure that the slight misalignment of the camera in the yaw direction generates a significant error (the graph shows the yaw error in the interval of ± 0.5 degrees). On the other hand, the effects of the roll and pitch errors are negligible compared to the discretization error (Figure 3 b)) shows the comparison of the roll and discretization error). Since the stereo setup for smoke detection will be working with relatively large distances we can discern the roll and pitch errors that have an insignificant effect on the accumulated error when compared to the discretization error.

For example, given a stereo setup with baseline $B = 2\,\mathrm{m}$, the focal length of both cameras $f = 4.655\,\mathrm{mm}$, pixel dimensions in the image plane $x_{pix} = 0.0022\,\mathrm{mm}$, where the observed object is located at distance of $Z = 500\,\mathrm{m}$, with a shift in the x dimension of $X2 = 7.5\,\mathrm{m}$ relative to the second camera. In case that the detection system requires the precision that would allow the depth estimation error to be no more than $100\,\mathrm{m}$ for this distance, we can calculate the maximum allowed error from different sources. Using the Equation (6) we can calculate that the maximum

discretization error for the given distance would be $Z_{disc\_err} = 66.9\,\text{m}$, so the rest of the error from other sources would be $33.1\,\text{m}$. It is important to emphasize that this is the maximum discretization error or the worst case scenario for the given distance. If we assume that each of the remaining types of errors (roll, pitch and yaw error) contribute evenly to the rest of the error it would mean that the maximum error from each of these sources would be less than $11.03\,\text{m}$. This means that the alignment accuracy for the given setup for specific rotations would be $\alpha < 6.19°$ (pitch error), $\beta < 0.0051°$ (yaw error) and $\theta < 6.21°$ (roll error). We omit the contribution of the lens distortion error for practical reasons.

From this example we can understand which calibration alignment types require more attention when constructing a stereo setup for smoke detection. Since we cannot influence the discretization error (except by widening the baseline or increasing the resolution of the image), the main key points are to ensure a thorough calibration of the system, and especially a very precise yaw orientation of the both cameras since it has a significant impact on the overall error.

4 IMPROVEMENT OF STANDARD DETECTION APPROACHES USING DEPTH INFORMATION

Standard visual smoke detection systems are usually equipped with a single rotating camera or a setup of several cameras pointed in different directions covering the 360° area around the detection post. The images acquired from the camera are analyzed in order to detect potential occurrence of smoke in the scene. In Section 2 we have covered the main phases of smoke detection process that are common to most smoke detection systems. However, using this kind of setup it is not possible to reliably estimate the distance of the detected phenomena from the camera. This additional information could be very useful in several phases of detection and help to improve the overall reliability of the system.

There are two main phases that could be improved using the information about the distance of the detected phenomena: region analysis phase and dynamics analysis phase. As covered in Section 2, the region analysis phase deals with the analysis of the candidate regions based on different smoke characteristics such as color, texture, shape and size. In order to eliminate noise from the detection process, most of the methods define a minimal size threshold for the detected region. This size is often expressed in the number of pixels that constitute the region. However, the number of pixels in the region does not provide the actual information about the physical size of the detected object. A group of pixels could represent a small object close to the camera or a large object at a great distance. One aspect of the improvement using stereo vision is to provide this additional information about the actual distance of detected objects. With this information it is possible to estimate the actual physical size of the detected objects. Objects or phenomena that do not satisfy minimal size constrains can now be eliminated from the detection process. This leads to a lower number of false alarms and a more reliable and robust detection process.

Another detection phase that could be improved is the dynamics analysis phase. This phase deals with the behavior of the candidate regions over time, where the regions are tracked over a certain period to verify that they exhibit smoke-like behavior. One of the main aspects of smoke dynamics is the growth rate of smoke regions in the image. In the incipient phase of wildfire, smoke gradually appears in the scene and continues to grow until reaching the full size. The growth rate and smoke size depend on many different parameters such as the fuel type, moisture content, wind speed and wind direction. However, it is possible to establish certain growth thresholds based on empirical data. By estimating actual physical size of the objects it is possible to eliminate regions in the image that exhibit very rapid growth that is much faster than the natural expansion of wildfire smoke. Eliminating such regions affects the performance of the system, by improving the reliability of the detection and reducing the number of false alarms.
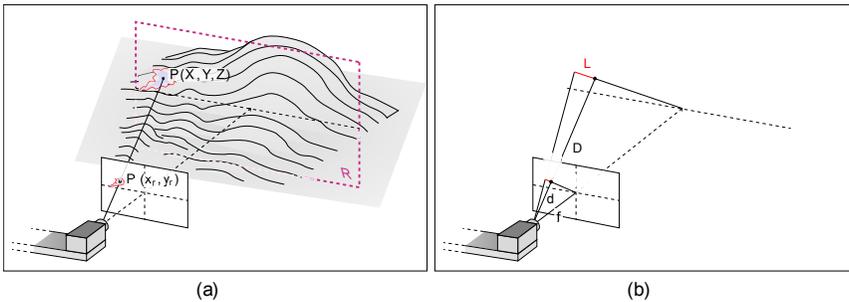

Figure 4. Calculating pixel coverage using: a) the relation between the point $P(X, Y, Z)$ in the real world and $P'(x_r, y_r)$ in the image plane, and b) the relation between shift L in the real world and the shift ℓ in the image plane

As already mentioned, due to the perspective of the camera, a group of pixels visible in the camera's image plane can represent a small object close to the camera, or a large object at a greater distance. Accordingly, we believe that it is better to express the size of the detected objects in standard units of measurement (such as meters) rather then in number of pixels.

Figure 4 illustrates the method used for estimating physical sizes of the objects visible in the scene. Using a stereo vision system introduced in previous sections, we can determine the depth (Z), as well as the actual distance (D) of the detected phenomena from the camera. Please recall, in this scenario the focal point of the right camera is chosen as the origin of the three-dimensional world coordinate system, although the system could easily be adapted to operate using the left camera. Also, as shown in Figure 1, the origin of the image coordinate system is positioned in the center of the camera image plane.

Let us assume that the detected phenomena is located at point $P(X, Y, Z)$ in the real world, and that point $P'(x_r, y_r)$ represents the projection of the point P onto the image plane of the (right) camera. We propose that the size of the de-

tected phenomena be estimated from its projection onto the plane that has the following characteristics: it and the detected phenomena are equal distance from the (right) camera, and it is parallel to the (right) camera's image plane (plane R in Figure 4 a)).

Before proceeding to the actual estimation of the physical size of the detected phenomena, let us make a remark: each pixel in the image also represents a space in the real world. This space can also be projected onto the described plane R, and therefore, the size of the space visible in the pixel can also be estimated.

Figure 4 b) demonstrates the method used to estimate the size of the physical space (visible inside a single pixel) projected onto the plane R that is parallel to the camera image plane. Let f represent the focal length of the (right) camera and d the distance from the focal point of the camera to the point $P'(x_r, y_r)$.

Let us first denote with ℓ the width of a single pixel (on the camera image plane). As illustrated in Figure 4 b), ℓ also represents the length of a line segment that lies on the line going through both the point $P'(x_r, y_r)$ and the center of the image plane. This shift of the length ℓ towards the edge of the camera image plane has a corresponding shift in the plane R, and let us denote it with L.

Variables f, d, ℓ and L can be expressed either in number of pixels or in standard units of measurement. Therefore, before proceeding, let us make some remarks concerning notation: from now on, all variables indexed with $(\cdot)_p$ will be associated with distance expressed in number of pixels, while the variables indexed with $(\cdot)_m$ will be associated with distance expressed in standard units of measurement (in this case meters). Our goal is to find the value of L_m that we will use to estimate the size of the physical space visible inside a single pixel.

Focal length is a parameter that is often provided by the manufacturers of the used equipment, but can also be retrieved by camera calibration. If given in number of pixels, it can easily be converted to standard units of measurement:

$$f_m = \frac{f_p \cdot ccd_m}{w_p} \tag{12}$$

where ccd_m represents the CCD width expressed in standard units of measurement (often millimeters) and w_p represents the width of the image expressed in number of pixels.

It can be shown that the value of d_m representing the distance from the focal point of the camera to the point $P'(x_r, y_r)$ can be calculated as follows:

$$d_m = \frac{f_m}{f_p} \cdot \sqrt{x_r^2 + y_r^2 + f_p^2}. \tag{13}$$

Please recall, the stereo vision system provides us with the value of D_m, therefore we can calculate the value of the shift L_m as follows:

$$L_m = \frac{f_m}{f_p} \cdot \frac{D_m}{d_m}. \tag{14}$$

Finally, the length of the shift L_m on the plane R that corresponds to the shift of one pixel on the camera image plane, expressed in standard units of measurement, is given by Equation (15).

$$L_m = \frac{D_m}{\sqrt{x_r^2 + y_r^2 + f_p^2}}.$$ (15)

Let us make the assumption that the pixels are square, then L_m^2 represents the area of the projected space onto the plane R that is parallel to the camera image plane. In other words, L_m^2 represents the size of the space visible in a single pixel positioned at coordinates (x_r, y_r) in the image plane. From now on, let us denote L_m^2 as a pixel coverage area.

From Equation (15) it can easily be seen that the pixel coverage area and the distance from the camera are directly proportional, meaning that the distant objects will appear smaller in the image. Using pixel coverage areas of all the pixels representing the detected phenomena, we can estimate it's overall physical size. However, stereo vision system does require some non-negligible amount of time to calculate the distance D_m. Hence, for practical reasons, we calculate the pixel coverage area only for one pixel positioned at the center of the detected region. The overall physical size can, therefore, be approximated as the pixel coverage area of the chosen pixel multiplied by the overall number of pixels representing the detected phenomena. In this manner, we do not slow down significantly the actual smoke detection method, since distance estimation using stereo vision is performed only once for each detected region in the image.

Now that we can calculate the pixel coverage area for a pixel in the center of the region, we can estimate the actual area of the detected objects in the real world. The detection method holds the information about the number of pixels in the candidate region, and by using the pixel coverage area of the central pixel we can estimate the area of the entire region. This area refers to the the area in the plane parallel to the image plane intersecting the actual object in physical space as depicted in Figure 4 a).

4.1 Filtering of the Candidate Regions Based on the Estimated Size

The information about the smoke area can now be used in the detection phases described earlier. The candidate regions detected as smoke in the region analysis phase can be further analyzed based on their real-world sizes. The candidate regions that are below the minimum size are eliminated from the detection process. The main reason for doing this is to reduce the number of false alarms. In fact, one of the major causes of noise, and therefore false alarms, is a movement of small objects visible in the scene. One example could be the uncontrolled movement of vegetation (such as grass or tree branches) in the close vicinity to the camera. Although these objects often share some similar characteristics with the smoke, they could be easily

dismissed as false alarms if their size is accurately approximated using the method described above.

In our case, we use the minimal size threshold of $5\,\mathrm{m}^2$ based on offline measurements and analysis. We believe that the most of the movement caused by smaller objects in the scene can be eliminated using this threshold, whereas, given the nature of the smoke, actual smoke regions will relatively quickly exceed this threshold.

Nevertheless, one important factor in the estimation of actual smoke size is the possible error due to the different error sources described in Section 3.2. When dealing with large distances, the discretization error becomes predominant due to the quadratic term of Equation (6). Therefore, Figure 5 shows the pixel coverage area for one pixel with regard to the distance from the camera. The figure also shows the pixel coverage area with the maximum positive and negative discretization errors (when the estimated distance is actually larger or smaller than the real distance of the object due to discretization). Recall, the maximum positive discretization error is larger then the maximum negative error as stated in Equation (6) (for the positive error, the second term in the denominator has a negative prefix).

From Figure 5 it is obvious that the area estimation becomes unreliable at larger distances, so it is important to ensure that valid region candidates are not rejected in this process. As already explained, this phase is used to reject candidates that are smaller than the predefined threshold ($5\,\mathrm{m}^2$). Please note that the positive discretization error results in larger pixel coverage area, so in this case the valid candidate regions are not rejected. However, the positive discretization error results in a non-rejection of noise regions that are estimated to have a size above the threshold due to this error. On the other hand, the negative discretization error may result in the rejection of valid candidate regions due to the negative error in size estimation.

We propose reducing the predefined minimal size threshold from initial $5\,\mathrm{m}^2$ to only $1\,\mathrm{m}^2$ if the estimated distance of the candidate region is beyond a certain distance D_{max} after which the area estimation becomes unreliable. In other words, D_{max} is a distance after which we cannot properly estimate the real-world size of the candidate region due to the discretization error. By doing this, we maintain the accuracy of the original smoke detection method at distances larger than D_{max} (for the same level of detection sensitivity), while at shorter distances we achieve the elimination of false alarms induced by the phenomena located in the vicinity of the camera. The value for D_{max} can be arbitrarily set based on the level of precision required from the system. For this purpose, we introduce a user defined limit err_ϑ defined as the maximum allowable percentage error of smoke area inside a single pixel. The percentage error (err) can be calculated as follows:

$$err = \frac{estimated\ area - real\ area}{real\ area} * 100. \qquad (16)$$

As an example, for $B = 2\,\mathrm{m}$, $f = 4.655\,\mathrm{mm}$ and $x_{pix} = 0.0022\,\mathrm{mm}$ and a chosen value for the maximum allowable error $err_\vartheta = 21\%$, based on Equations (6),
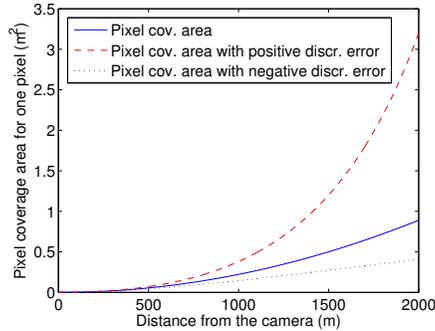
Figure 5. Pixel coverage area (m^2) for one pixel with regard to the distance from the camera (blue solid line), pixel coverage area with maximum positive discretization error for the given distance (red dashed line), and pixel coverage area with maximum negative discretization error for the given distance (black dotted line). The selected pixel has coordinates $(100, 100)$ referent to the center of the image, with $B = 2\,\text{m}$, $f = 4.655\,\text{mm}$ and $x_{pix} = 0.0022\,\text{mm}$.

(15) and (16), it is possible to calculate the value for the maximum distance with allowable error in front of the camera $D_{max} = 385\,\text{m}$.

In this way, the reliability of the pixel coverage estimation could be adjusted based on the needs of the specific detection method or system. If the percentage error increases over the user defined threshold, the minimum size threshold reduces to $1\,\text{m}^2$ in order to compensate for the possible discretization error. It is important to emphasize that the detection method is still capable of detecting smoke beyond the distance specified by D_{max} with the same accuracy as the original version. The reduction of minimum size threshold is to ensure that real smoke plumes are not discarded based on size due to discretization error. This implies that the detection of smoke is sill accurate, but the benefit gained by rejection of potential false alarms based on size reduces with distance.

Another phase where information about pixel coverage area can be useful for reducing false alarm rate and improving the reliability of the detection process is the dynamics analysis phase. The dynamics analysis phase is used to eliminate regions that do not exhibit smoke-like behavior. Smoke behavior is rather difficult to define or simulate since it depends on many different factors as described earlier in this section. However, overall smoke dynamics adhere to certain rules regarding the rate of spread. The measurements published in [30] show that the average smoke area rate of spread (μ) in the image plane is $16.04\,\text{m}^2/\text{s}$ with standard deviation (σ) $76.33\,\text{m}^2/\text{s}$ in the first 3 minutes after the occurrence. It is also stated that by fitting the data with t location-scale distribution it can be calculated that over 99 percent of smoke area change observations fall into the interval defined by $(\mu - 6\sigma, \mu + 6\sigma)$.

Using this information, and the information about the area of the candidate regions calculated using the pixel coverage area, it is now possible to reject those

regions with grow or shrink rates outside the allowed range. This consequently results in higher reliability and accuracy of the detection method or system.

As already mentioned, by reducing the false alarms rate of the existing smoke detection methods, it is possible to increase detection sensitivity, allowing the increase of both correct detections rate and the coverage area where the system can detect wildfires. In the following section we will present the evaluation process and methodology as well as the obtained results.

5 EVALUATION

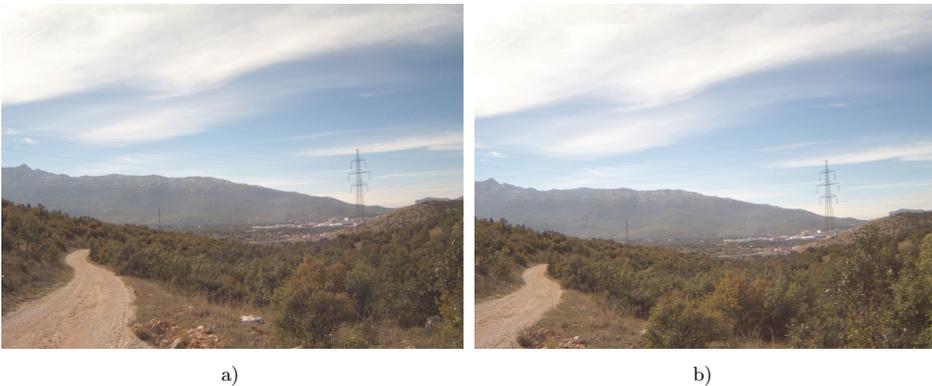a)   b)

Figure 6. An example of stereo images captured simultaneously by both cameras. Input image taken by: a) the left stereo camera, b) the right stereo camera.

The proposed improvement based on stereo vision was implemented into three existing smoke detection methods. Every method was tested with and without this improvement on a database consisting of 18 856 images. Since all the video sequences were recorded using stereo cameras, the database actually consisted of altogether 37 712 images. These images were extracted from video sequences every 1 s, meaning that the total time span of the footage is approximately 5 hours and 15 minutes. In 9 198 images (approximately 2 hours and 30 minutes) smoke is visible in the image in its various forms. The remaining 9 658 images include various other phenomena that could induce false alarms (e.g., vegetation movement caused by wind, shadowing by clouds, changes in lightning conditions during sunrise or sunset). For a quality evaluation of the proposed improvement based on stereo vision, it is important to use both types of images in order to properly examine all the quality measures of the system. In Figure 6 we show a pair of stereo images that are a part of our database.

Video sequences were recorded on various locations and under different weather conditions to ensure diversity of the scenes. Smoke phenomena captured on video

were located at different distances from the cameras to ensure that the evaluation takes into account possible errors of the stereo vision system (discretization and camera alignment errors).

Videos were recorded using two "`Elphel NC353L`" video cameras that were mutually synchronized using external synchronization cable (GPIO). This enabled the synchronization precision up to 1μ s and simultaneous capturing of the images on both cameras. The focal length of the left camera was 4.657 mm, while the focal length of the right camera was 4.778 mm. In order to be able to use Equation (3) for our calculations, we first had to compensate this difference in focal lengths (as well as differences in principal point coordinates and skew and distortion parameters) by normalizing the pixel coordinates used as the input for the aforementioned equation. The images were captured in various resolutions. Both cameras were securely mounted on the ends of a wide bar, making the length of the baseline precisely 2 m. The maximum allowable percentage error of smoke area inside a pixel for the stereo system is set to $err_\vartheta = 21\%$. In Figure 7 a) we can see a close-up of the built stereo vision system, while in Figure 7 b) we show the system operating in the natural environment with predominant vegetation.

a)   b)

Figure 7. The stereo system used for evaluation, a) in laboratory environment, and b) in natural environment

As mentioned above, three existing smoke detection methods were used for the evaluation. First, each method was tested without any stereo modifications on all 18 856 images taken by the right camera. Second, improved methods were tested on the images of both cameras. In the following, we give a brief explanation of the methods used for evaluation.

We have implemented *Method 1* presented in [5], which is a wavelet based real-time smoke detection method. The method consists of several detection phases. In the first phase, the moving pixels in the image are detected using a background estimation method. In the next step, the high frequency content of the image is analyzed in order to detect blurring due to the possible occurrence of smoke. Appearance of smoke gradually reduces the sharpness of the edges in the region until the region is completely covered by smoke. The region is analyzed in order to detect a decrease in local wavelet energy which would suggest presence of smoke. In the

following phase, the regions are checked for decrease in the U and V channels. The appearance of smoke in the region results in the decrease in the chrominance level when compared to the estimated background. The next phase of the algorithm deals with analysis of the flickering effect that appears on the edges of the smoke contour. The analysis is carried out using temporal wavelet transforms. It is important to emphasize that this effect is noticeable in the short range smoke detection, and is not significant when the smoke is located at larger distances from the camera. The final phase of the algorithm examines the shape of the detected region in order to determine its convexity since general wildfire smoke tends to have a rather convex shape. In the case when the criteria from all the phases are satisfied, the algorithm raises an alarm.

The second method that we have implemented is the method presented in [18] denoted as *Method 2*. This method does not operate on the whole captured image, but the image is rather divided into blocks, or bins, which represent the smallest units for the detection process. The blue channel of each bin is observed over time and compared against the signal range, i.e. the difference between the maximum and the minimum bin value. The blue channel is selected since it exhibits greater sensitivity to smoke appearance then the other channels. In the case the difference between the current bin value and the referent bin value exceeds the percentage threshold based on signal range, the bins are considered as the candidate bins. The algorithm tracks the bins over a certain period of time in order to confirm the potential detection before raising an alarm. To produce the final alarm, there has to be a certain minimal number of candidate regions present over a set time period. Additionally smoke has to make a gradual appearance on the scene, so the regions that exhibit growth between two consecutive captured images that is larger than the maximum permitted growth are dismissed. Another possible filter for rejection of false alarms is the maximum permitted number of 8-connected components in the image. In the case all the conditions are satisfied over a predefined period of time, the system consequently raises an alarm.

The third method is the method presented in [31] denoted as *Method 3*. This method consists of several different phases or stages of detection. The first phase is the image segmentation and classification phase where the different classes such as water and sky are used for the elimination of possible false alarms. Next, a motion detection phase is used to detect only the moving regions in the image, and thus reducing the amount of data that requires further processing. The following phase performs chromatic analysis where the current chromatic values are compared to the referent smoke-color values. The next step is the texture analysis phase where the regions are analyzed based on wavelet information. In the case of smoke appearance in the scene the region texture should change and there should be a loss in the high frequency range due to the blurring caused by smoke. In the following phase, the dynamic aspect of the candidate regions is examined. The regions that do not exhibit smoke-like behavior over a predefined time period are eliminated from the detection process. Finally, in case the candidate regions are confirmed in each detection phase, they are considered to be smoke, and the alarm is raised by

the algorithm. The example detection images for all methods are shown in Figure 8.
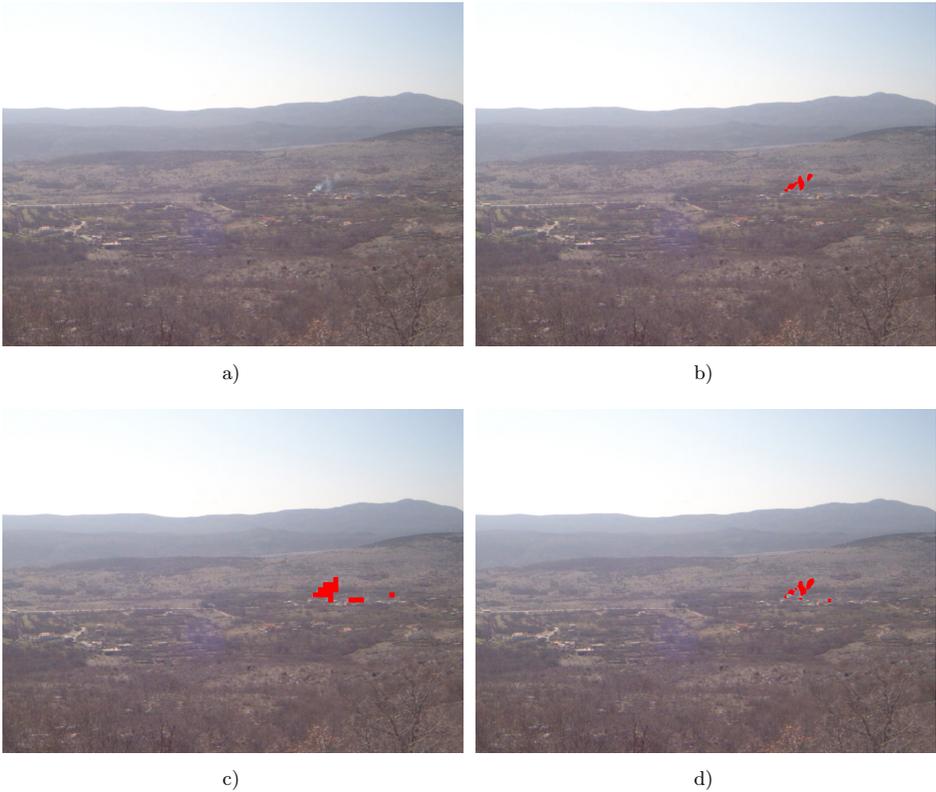

Figure 8. a) Input image with visible smoke, taken by the right stereo camera, b) image with smoke detected by *Method 1*, c) image with smoke detected by *Method 2*, d) image with smoke detected by *Method 3*

For each method, a stereo version was implemented based on the features presented in previous sections (the minimum size and overall dynamics of the candidate regions are examined to verify that they exhibit characteristic features of smoke). Here, we shall discuss the evaluation of each method and compare the results with the improved versions. We will use the evaluation measures for visual smoke detection systems presented in [32]. The evaluation is divided into two main evaluation aspects, global and local evaluation. Global measures evaluate algorithm performance based on the results where the elemental evaluation units are images. The evaluation is based on the algorithm output, where the algorithm decides whether smoke is present in the image. We can use four measures describing different aspect of detection quality: measure correct detections, also known as recall or true positive

rate (TPR), specificity or true negative rate (TNR), false positive rate (FPR) and false negative rate (FNR). The measures are defined as:

$$TPR = \frac{TP}{TP + FN},$$ (17)

$$TNR = \frac{TN}{TN + FP},$$ (18)

$$FPR = \frac{FP}{TN + FP},$$ (19)

$$FNR = \frac{FN}{TP + FN}$$ (20)

where TP denotes the number of true positive detections, FN represents the number of false negative detections, TN represents the number of true negative detections, and FP represents the number of false positive detections.

The results for global measures for all methods are presented in Table 1.

	TPR	TNR	FPR	FNR
Method 1 – standard	0.6601	0.8701	0.1299	0.3399
Method 1 – stereo	0.6677	0.9622	0.0378	0.3323
Method 2 – standard	0.4322	0.9994	0.0006	0.5678
Method 2 – stereo	0.4240	0.9996	0.0004	0.5760
Method 3 – standard	0.6954	0.8884	0.1116	0.3046
Method 3 – stereo	0.7024	0.9358	0.0642	0.2976

Table 1. Results for global measures for all methods

The results are obtained from all evaluation sequences. The stereo versions of the evaluated methods show a general improvement for most evaluation measures, especially a decrease in false alarms with similar or improved correct detections. The decrease in false alarms is a result of region size estimation based on the stereo distance calculation. Majority of false alarm sources are eliminated with this process, such as movements of the vegetation or similar phenomena in the close proximity of the camera. Most of the smoke detection methods have a set of tunable parameters which define the sensitivity of detection with respect to the dynamics in the environment. Increase in the algorithm sensitivity results in a more prompt and precise detection of actual smoke in the scene, however, it also results in a general increase in false alarms. Reducing the sensitivity of the algorithm, on the other hand, decreases false alarm rate, but increases the risk of missed detections. In the case when a tool for elimination of false alarms is introduced, such as stereo distance estimation, it is possible to increase the sensitivity of the algorithm. The possible increase in false alarms is in this case compensated with the false alarms elimination process. That is the reason why some of the results also show the increase in correct detections (recall), as it is the case with *Method 1* and *Method 3*. The exception is a slight

decrease in correct detections for *Method 2* due to the fact that the standard version
has a very low false alarm rate to begin with. Please note that by increasing the
detection sensitivity we also increase the coverage area, i.e., the maximal distance
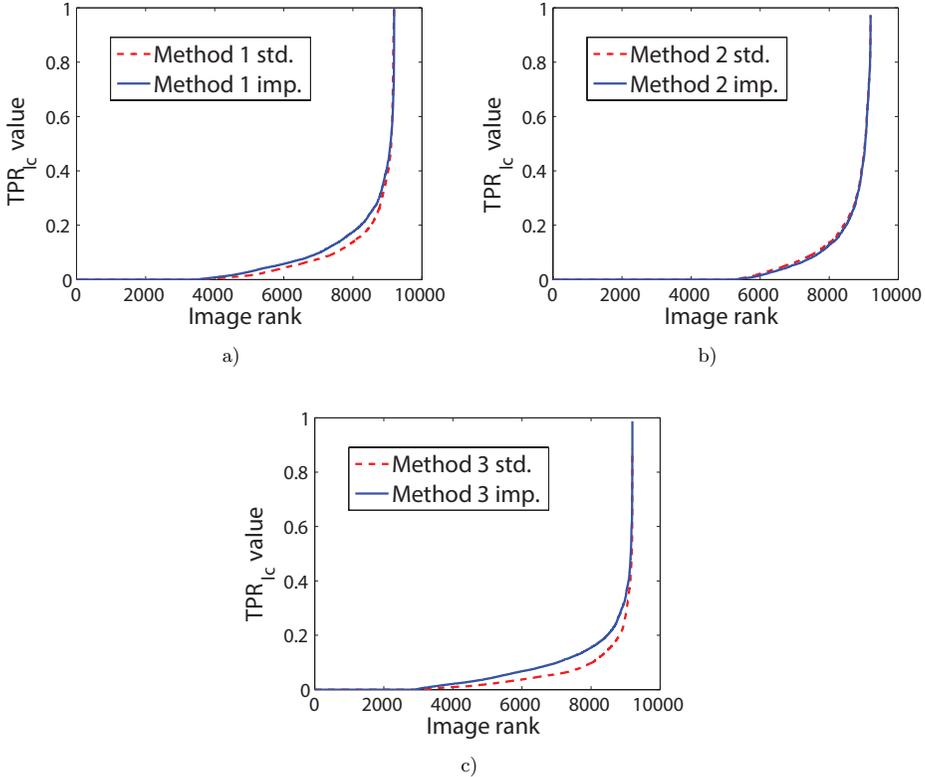at which the method can detect wildfire.

a)

b)

c)

Figure 9. Observer quality graphs for TPR_{lc} measure: a) *Method 1*, b) *Method 2*, and c)
Method 3

Another type of evaluation is performed using local measures. Local measures
are based on the results where the smallest units of detection are individual pixels.
Global measures are focused on whether the smoke is detected in the image or
not, while the local measures are focused on whether the location of the smoke in
the image is correct or not. Generally, when evaluating smoke detection systems,
the most important fact is that the alarm is raised with the occurrence of smoke,
with as low as possible false alarm rate. The location of the smoke in the image
is also important, however, the global performance is the primary criterion. We
have performed the evaluation on the local scale for measures described earlier, for
all detection methods. The results are presented in the form of observer quality
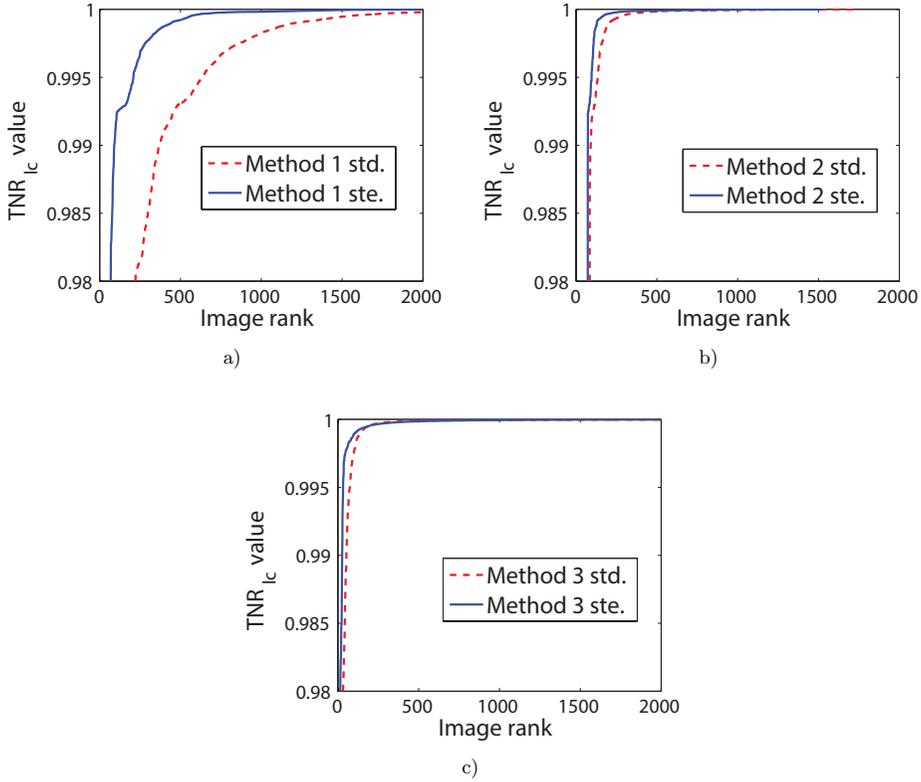
Figure 10. Observer quality graphs for TNR_{lc} measure: a) *Method 1*, b) *Method 2*, and c) *Method 3*

graphs introduced in [32]. The graphs show the value of the specific measure for all the images in the collection, sorted according to the increasing measure values. In the graphs, the y axis represents the value of the specific measure, while the x axis represents the ordinal number of the image in the sorted sequence, or the image rank. Figures 9–12 show the observer quality graphs for local measures of correct detections or true positive rate (TPR_{lc}), true negative rate (TNR_{lc}), false positive rate (FPR_{lc}) and false negative rate (FNR_{lc}), respectively.

The local correct detections measure (TPR_{lc}) deals with correctly detected smoke pixels within the images in the collection. The stereo versions of the detection methods show same or better local accuracy of correct detections when compared to the standard variants as seen in Figure 9. When comparing the results for local specificity measures (TNR_{lc}), all of the stereo methods exhibit an increase in valid local rejections when compared to the standard versions as shown in Figure 10. This effect is the result of elimination of certain categories of false alarms based
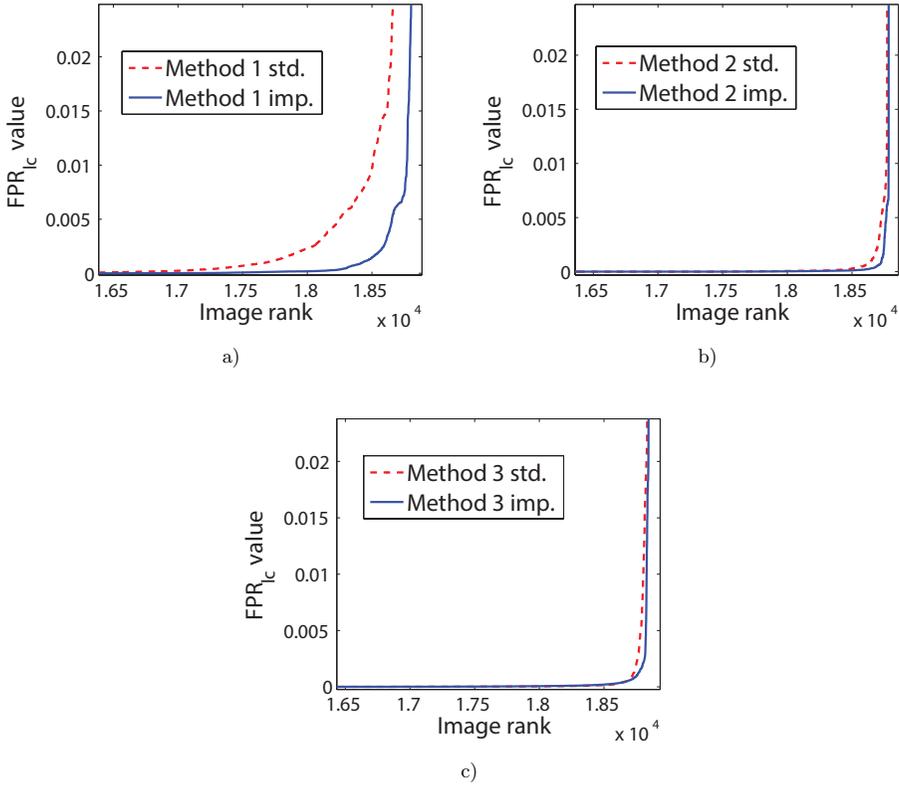
Figure 11. Observer quality graphs for FPR_{lc} measure: a) *Method 1*, b) *Method 2*, and c) *Method 3*

on stereo distance estimation. Similarly, this also affects the results concerning local false alarms where all stereo methods exhibit a drop in local false alarms as shown in Figure 11. Finally, the measure for local false negative rate (FNR_{lc}) for stereo versions remains the same or exhibits a drop in missed detections as shown in Figure 12.

Another measure that can be applied to the evaluation of smoke detection algorithms on the local scale is the Matthews correlation coefficient [33]. The Matthews correlation coefficient (mcc) is defined by:

$$mcc = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{21}$$

and is generally used as a quality measure for binary classifications. This measure takes into account all the detection cases (TP, TN, FP, FN) in a balanced manner
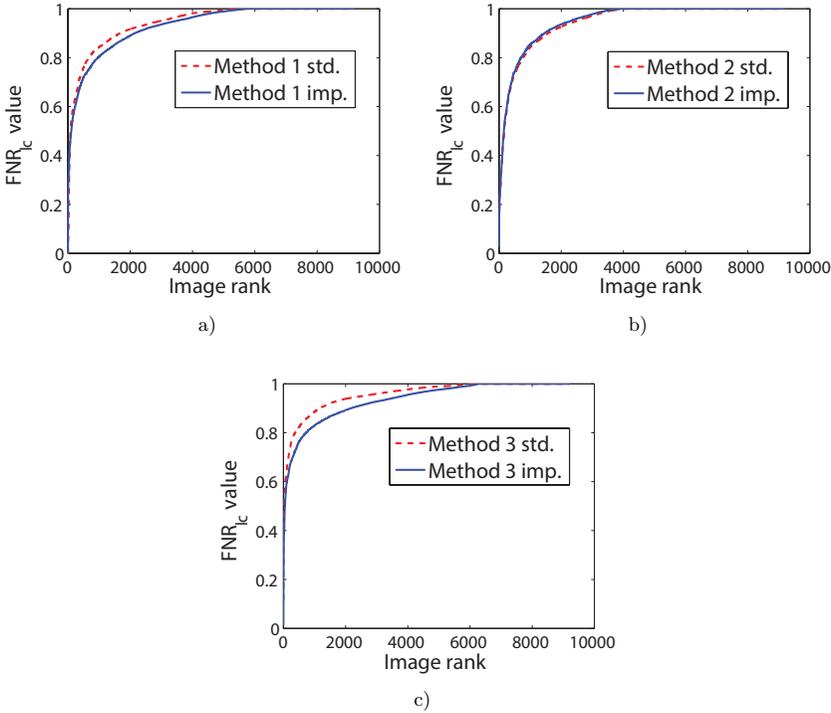
Figure 12. Observer quality graphs for FNR_{lc} measure: a) *Method 1*, b) *Method 2*, and c) *Method 3*

so it can be used even when the classes are of very different sizes. The results for the *mcc* measure for all methods are shown in Figure 13.

The *mcc* measure result values are in the interval $[-1, 1]$. The results for the stereo versions of the detection methods show the same or increased value when compared to the standard versions, which indicates an improvement in the overall local detection quality.

6 CONCLUSIONS

In this paper we proposed a novel approach to smoke detection by introducing stereo vision to the detection process. Standard smoke detection techniques available in literature, can be extended in a way to include stereo vision techniques allowing us to estimate the real world sizes of the detected regions and express them in standard units of measurement.

This provides the detection system with additional information that could be used to improve the reliability of the detection process. Using this information, the
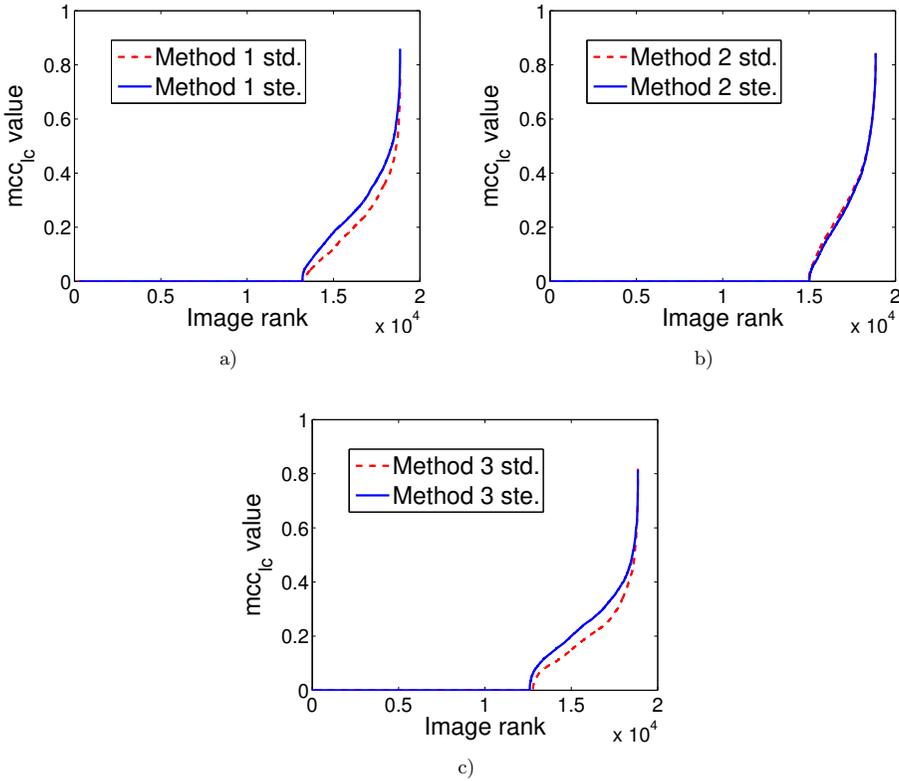
Figure 13. Observer quality graphs for mcc_{lc} measure: a) *Method 1*, b) *Method 2*, and c) *Method 3*

minimum size and overall dynamics of the candidate regions are examined to verify that they exhibit characteristic features of smoke. In this way, it is possible to determine the phenomena that cause false alarms due to their resemblance to smoke and eliminate them from the detection based on size and dynamics constraints. As an example, minimum size constraint is used for the elimination of false alarms induced by movement of small objects in the scene (such as movement of branches and grass in the close vicinity of the camera). Additionally, smoke dynamics are analyzed to verify that regions adhere to certain rules regarding the rate of spread, leading to the elimination of false alarms induced by phenomena that have significantly different dynamics characteristics (such as shadowing by clouds).

Since smoke detection implies estimating relatively large distances, we have proposed a wide-baseline stereo vision system. We have analyzed possible errors that might affect the accuracy of the stereo system as well as their impact on the actual depth estimation. These errors include discretization error, that is a result of

an approximation due to the discrete nature of the imaging system, and camera alignment errors, more specifically the errors due to roll, pitch and yaw of one of the stereo cameras.

We have evaluated three existing smoke detection methods and compared them to the newly implemented stereo versions. The results based on global and local evaluation measures have shown improved overall performance, especially in false alarms for all tested methods. Moreover, by increasing the detection sensitivity we not only achieved a lower false alarms rate, but also either maintained or increased the number of correct detections and the coverage area of the existing methods. This shows that this improvement based on stereo vision could be used as a welcome addition to the standard detection methods in terms of system reliability and could also be taken into account when designing future smoke detection systems, especially if intended for areas with a high risk of wildfires.

REFERENCES

[1] DUKUZUMUREMYI, J. P.—ZOU, B.—HANYURWIMFURA, D.: A Novel Algorithm for Fire/Smoke Detection Based on Computer-Vision. International Journal of Hybrid Information Technology, Vol. 7, 2014, No. 3, pp. 143–154, doi: 10.14257/ijhit.2014.7.3.15.

[2] MORIMITSU, N.: Image Processing Apparatus. US Patent Application US20120133739A1, May 31, 2012.

[3] BUGARIĆ, M.—JAKOVČEVIĆ, T.—STIPANIČEV, D.: Adaptive Estimation of Visual Smoke Detection Parameters Based on Spatial Data and Fire Risk Index. Computer Vision and Image Understanding, Vol. 118, 2014, pp. 184–196, doi: 10.1016/j.cviu.2013.10.003.

[4] CAPPELLINI, V.—MATTII, L.—MECOCCI, A.: An Intelligent System for Automatic Fire Detection in Forests. Third International Conference on Image Processing and Its Applications, July 1989, pp. 563–570, doi: 10.1007/3-540-51815-0_67.

[5] TOREYIN, B.—DEDEOGLU, Y.—CETIN, A.: Wavelet Based Real-Time Smoke Detection in Video. Proceedings of the 13th European Signal Processing Conference (EUSIPCO 2015), Vol. 20, 2005, pp. 1–4.

[6] XU, Z.—XU, J.: Automatic Fire Smoke Detection Based on Image Visual Features. Proceedings of the International Conference on Computational Intelligence and Security Workshops (CISW 2007), 2007, pp. 316–319, doi: 10.1109/CISW.2007.4425500.

[7] KIM, D.—WANG, Y.-F.: Smoke Detection in Video. Proceedings of WRI World Congress on Computer Science and Information Engineering, 2009, pp. 759–763, doi: 10.1109/CSIE.2009.494.

[8] OCHOA-BRITO, A.—MILLAN-GARCIA, L.—SANCHEZ-PEREZ, G.—TOSCANO-MEDINA, K.—NAKANO-MIYATAKE, M.: Improvement of a Video Smoke Detection Based on Accumulative Motion Orientation Model. Proceedings of the Electronics, Robotics and Automotive Mechanics Conference (CERMA 2011), 2011, pp. 126–130.

[9] Ho, C.—Kuo, T.: Real-Time Video-Based Fire Smoke Detection System. Proceedings of the International Conference on Advanced Intelligent Mechatronics, 2009, pp. 1845–1850.

[10] Ma, L.—Wu, K.—Zhu, L.: Fire Smoke Detection in Video Images Using Kalman Filter and Gaussian Mixture Color Model. Proceedings of the International Conference on Artificial Intelligence and Computational Intelligence (AICI), Vol. 1, 2010, pp. 484–487, doi: 10.1109/AICI.2010.107.

[11] Piccinini, P.—Calderara, S.—Cucchiara, R.: Reliable Smoke Detection in the Domains of Image Energy and Color. Proceedings of the 15th IEEE International Conference on Image Processing (ICIP 2008), 2008, pp. 1376–1379, doi: 10.1109/ICIP.2008.4712020.

[12] Krstinić, D.—Stipaničev, D.—Jakovčević, T.: Histogram-Based Smoke Segmentation in Forest Fire Detection System. Information and Technology Control, Vol. 38, 2009, No. 3, pp. 237–244.

[13] Anton, M.—Olga, K.: Real-Time Smoke Detection in Video Sequences: Combined Approach. In: Maji, P., Ghosh, A., Murty, M. N., Ghosh, K., Pal, S. K. (Eds.): Pattern Recognition and Machine Intelligence (PReMI 2013). Springer Berlin, Heidelberg, Lecture Notes in Computer Science, Vol. 8251, 2013, pp. 445–450.

[14] Busch, A.—Boles, W. W.: Texture Classification Using Multiple Wavelet Analysis. Proceedings of the Sixth Digital Image Computing – Techniques and Applications Conference (DICTA 2002), 2002, pp. 341–345.

[15] Kara, B.—Watsuji, N.: Using Wavelets for Texture Classification. Proceedings of the International Conference on Signal Processing, 2003, pp. 920–924.

[16] Yu, C.—Zhang, Y.—Fang, J.—Wang, J.: Texture Analysis of Smoke for Real-Time Fire Detection. Proceedings of the Second International Workshop on Computer Science and Engineering (WCSE '09), 2009, pp. 511–515.

[17] Rafiee, A.—Dianat, R.—Jamshidi, M.—Tavakoli, R.—Abbaspour, S.: Fire and Smoke Detection Using Wavelet Analysis and Disorder Characteristics. Proceedings of the 3rd International Conference on Computer Research and Development (ICCRD), 2011, pp. 262–265, doi: 10.1109/ICCRD.2011.5764295.

[18] Fernández-Berni, J.—Carmona-Galán, R.—Carranza-González, L.: A Vision-Based Monitoring System for Very Early Automatic Detection of Forest Fires. Proceedings of the First International Conference on Modelling, Monitoring and Management of Forest Fires (FIVA 2008), 2008, pp. 161–170, doi: 10.2495/FIVA080171.

[19] Chen, J.—You, Y.—Peng, Q.: Dynamic Analysis for Video Based Smoke Detection. International Journal of Computer Science Issues, Vol. 10, 2013, No. 2, pp. 298–304.

[20] Cui, Y.—Dong, H.—Zhou, E.: An Early Fire Detection Method Based on Smoke Texture Analysis and Discrimination. Proceedings of the Congress on Image and Signal Processing (CISP '08), 2008, pp. 95–99, doi: 10.1109/CISP.2008.397.

[21] Ko, B. C.—Kwak, J.-Y.—Nam, J.-Y.: Wildfire Smoke Detection Using Temporospatial Features and Random Forest Classifiers. Optical Engineering, Vol. 51, 2012, Art. No. 017208, doi: 10.1117/1.OE.51.1.017208.

[22] YANG, J.—CHEN, F.—ZHANG, W.: Visual-Based Smoke Detection Using Support Vector Machine. Proceedings of the Fourth International Conference on Natural Computation (ICNC '08), 2008, pp. 301–305, doi: 10.1109/ICNC.2008.219.

[23] STIPANIČEV, D.—ŠERIĆ, L.—BRAOVIĆ, M.—KRSTINIĆ, D.—JAKOVČEVIĆ, T.—ŠTULA, M.—BUGARIĆ, M.—MARAS, J.: Vision Based Wildfire and Natural Risk Observers. Proceedings of the 3rd International Conference on Image Processing Theory, Tools and Applications (IPTA), OS1: Special Session on Image Processing for Natural Risks (IPNR), 2012.

[24] MROVLJE, J.—VRANČIĆ, D.: Distance Measuring Based on Stereoscopic Pictures. 9th International Ph.D. Workshop on Systems and Control: Young Generation Viewpoint, 2008.

[25] GALLUP, D.—FRAHM, J.-M.—MORDOHAI, P.—POLLEFEYS, M.: Variable Baseline/Resolution Stereo. 26th IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2008, pp. 1–8.

[26] VRANCIC, D.—SMITH, S. L.: Permanent Synchronization of Camcorders via LANC Protocol. In: Woods, A. J., Bolas, M. T., McDowall, I. E., Dodgson, N. A., Merritt, J. O. (Eds.): Stereoscopic Displays and Virtual Reality Systems XIII. SPIE Proceedings, Vol. 6055, 2006, pp. 165–176.

[27] LEWIS, J. P.: Fast Normalized Cross-Correlation. Vision Interface, Vol. 10, 1995, No. 1, pp. 120–123.

[28] SAHABI, H.—BASU, A.: Analysis of Error in Depth Perception with Vergence and Spatially Varying Sensing. Computer Vision and Image Understanding, Vol. 63, 1996, No. 3, pp. 447–461.

[29] ZHAO, W.—NANDHAKUMAR, N.: Effects of Camera Alignment Errors on Stereoscopic Depth Estimates. Pattern Recognition, Vol. 29, 1996, No. 12, pp. 2115–2126.

[30] BUGARIĆ, M.—JAKOVČEVIĆ, T.—STIPANIČEV, D.: Computer Vision Based Measurement of Wildfire Smoke Dynamics. Advances in Electrical and Computer Engineering, Vol. 15, 2015, No. 1, pp. 55–62.

[31] JAKOVČEVIĆ, T.—STIPANIČEV, D.—KRSTINIĆ, D.: Visual Spatial-Context Based Wildfire Smoke Sensor. Machine Vision and Applications, Vol. 24, 2013, No. 4, pp. 707–719, doi: 10.1007/s00138-012-0481-x.

[32] JAKOVČEVIĆ, T.—ŠERIĆ, L.—STIPANIČEV, D.—KRSTINIĆ, D.: Wildfire Smoke-Detection Algorithms Evaluation. In: Viegas, D. X. (Ed.): Proceedings of the VI International Conference on Forest Fire Research, 2010, pp. 1–12.

[33] MATTHEWS, B. W.: Comparison of the Predicted and Observed Secondary Structure of T4 Phage Lysozyme. Biochimica et Biophysica Acta (BBA) – Protein Structure, Vol. 405, 1975, No. 2, pp. 442–451, doi: 10.1016/0005-2795(75)90109-9.

Toni Jakovčević graduated in computer science from the University of Split, Croatia, and afterwards received his Ph.D. in 2011. His research interests include artificial intelligence, image analysis and programming languages. Currently, as Assistant Professor, he conducts research, development and teaching in the University of Split.

Marin Bugarić holds the position of a young researcher at the University of Split, Croatia. He received his Ph.D. in 2013. His research interests include early forest fire detection systems based on visual smoke detection, GIS applications, augmented reality systems, forest fire risk indexes, usability, design and analysis of the secure authentication protocols.

Darko Stipaničev is Professor of computer science and automatic control at the University of Split. His research interests include complex systems modeling and control, intelligent systems analyses and design, digital image analyses, advanced Internet technologies, and recently, the application of ICT in environmental protection, particularly wildfires prevention and management, including vision-based wildfires smoke detection.