A Camera Uncertainty Model for Collaborative Visual Sensor Network Applications

Christos Kyrkou, Eftychios Christoforou, Theocharis Theocharides, Christos Panayiotou, Marios Polycarpou
Department of Electrical and Computer Engineering
KIOS Research Center for Intelligent Systems and Networks
University of Cyprus
{kyrkou.christos, e.christoforou, ttheocharides, christosp, mpolycar}@ucy.ac.cy

ABSTRACT

Visual Sensor Networks (VSNs) exploit the processing and communication capabilities of modern smart cameras to handle a variety of applications such as security and surveillance, industrial monitoring, and critical infrastructure protection. The performance of VSNs can be severely degraded because of errors in the detection module. As a result, the performance of the higher-level application such as activity recognition, tracking, etc., also suffers due to the fact that in most cases the decision making process in VSNs assumes ideal detection capabilities for the cameras. Realizing that it is necessary to introduce robustness in the decision process this paper presents results towards uncertainty-aware VSNs. Specifically, we introduce a flexible uncertainty model that can be used to study the behaviour of missed detections in a camera network. We also show how to utilize the model to develop uncertainty-aware coordination and decision making solutions to improve the efficiency of VSNs. Our experimental results in an active vision application indicate that the proposed solution is able to improve the robustness and reliability of VSNs.

Keywords

Active Vision; Self-Configuring Camera Networks; Visual Sensor Networks; Camera Sensing Model; Smart Camera

1. INTRODUCTION

Visual Sensor Networks (VSNs) consist of networked cameras that can communicate and perform multiple vision tasks (activity recognition, tracking, etc.) while observing a scene [1]. Recently, emerging VSNs offer advanced sensing and processing capabilities and collaboration capabilities that facilitate the development of a wide range of applications ranging from security and surveillance, automated transportation systems, personalized healthcare, industrial monitoring and augmented reality [2], [3]. The performance of these applications relates directly to the detection module capabilities of the VSN as missed detections can cause unpredictable behaviour and compromise the decision making process. Hence, it is of key importance to develop models, algorithms, and systems that take into consideration different uncertainties in VSNs and use them to increase the robustness of the application.

The majority of existing works assume that VSNs operate under perfect conditions, and do not take into account the possibility that a camera may not detect an object even if it is visible by the visual sensor. Realistically, even cameras featuring sophisticated visual sensors and on-board processors for decision-making, are inherently error prone due to the probabilistic nature of the detection algorithms, and so may provide wrong decisions for different reasons. For example, if a camera fails to detect a specific target, due to a slight change in viewpoint or insufficient representative pixel resolution, then that camera would falsely report that the target is not present, while it moves inside its Field-of-View (FoV). This becomes more apparent for low-cost camera systems [5] which do not have the resources to efficiently run state-of-the-art detection algorithms and so either lower the resolution or run less demanding algorithms both of which compromise detection performance. Only a few works have considered issues that relate with unreliability in the context of VSNs. For example, the work in [4] investigates the impact of errors in the horizontal orientation (i.e., pan) of cameras during target tracking, due to initial calibration inaccuracies (modelled as Gaussian noise) and external effects that cause the camera orientation to take arbitrary values.

Motivated by the importance of dealing with uncertainties in VSN applications this paper presents an effort to introduce uncertainty awareness into VSNs to improve their efficiency. To this end, the contribution of this work is twofold. First, we propose a flexible uncertainty model based on detection-probability/confidence zones that can be used to study the impact of degrading detection accuracy in VSN applications. To the best of our knowledge, this is the first attempt to develop an uncertainty model for VSNs. Second, we utilize the model to characterize the detection capabilities of each camera and use it to improve the coordination and decision making in collaborative VSNs. We show the application of this confidence zone model in an active vision scenario using a network of Raspberry-Pi-based pan-tilt smart cameras, where cameras need to reconfigure in order to reduce the uncertainty by which they monitor a target.

The rest of this paper is structured as follows. Section II outlines some key areas of emerging research in VSNs. In Section III we introduce the network, sensing, target and uncertainty models and the underlying assumptions. Section IV presents the details of a decision-making mechanism, and a dynamic pan-tilt (PT) reconfiguration scheme that utilize the proposed confidence zone model. In Section V we experimentally evaluate the validity of the model and apply the dynamic reconfiguration scheme to an active vision problem where cameras collaboratively decide on how to adjust their parameters in order to meet the required confidence level. Finally, Section VI provides concluding remarks and discusses directions for future work.

2. RELATED WORK

Emerging research developments in VSNs have led to customized smart camera systems [5], and distributed vision algorithms (e.g., [6]); both of which have enabled new applications but simultaneously introduced new challenges. One such challenge concerns pan-tilt-zoom (PTZ) cameras and how to use them efficiently. Thus, there has been a lot of emerging research concerning PTZ cameras and networks [7], as well as dynamic network reconfiguration [8]. Improving the performance, image-based control, and automated tracking aspects of single PTZ cameras has been the subject of many works in the literature [9],[10],[11]; as this can also provide higher efficiency when considering a network of such cameras. Consequently, there has also been an increasing amount of research effort, towards improving various aspects of VSNs by utilizing information from multiple cameras. For example, [12],[13] deal with the problem of naivety in static VSNs where not all cameras observe all targets, but need to maintain a state estimate for each target. The work in [14] reconfigures a network of PTZ cameras to maintain the overall network coverage in order to compensate for changes in the viewpoint of cameras that opportunistically zoom into an area to take high-resolution images. However, in their majority, these works do not consider uncertainties in the camera detection modules and how this can degrade performance. As such, existing models and approaches assume that a visual sensor will always detect a target that is present in its FoV. In this paper, we present research towards the development of an uncertainty model that can describe the detection behaviour of a camera and can thus provide input to decision-making and dynamic configuration algorithms in order to improve their outcome.

3. MODEL DESCRIPTION

A visual sensor network is considered consisting of N_C static camera nodes C_j that belong to the set $\boldsymbol{\mathcal{C}} = \{C_1, C_2, \dots, C_{N_C}\}$ and N_T targets T_i to the set $\boldsymbol{\mathcal{T}} = \{T_1, T_2, \dots, T_{N_T}\}$ that need to be monitored. Furthermore, the camera can detect a target with varying probability based on its position, and viewpoint with regards to it. The assumptions and network model are described next.

3.1 Visual Sensor Model

Cameras in the network may be static, in which case their orientation and position does not change. In the case of an active network, the cameras have some degrees of freedom and can manipulate their point of view based on collective or local information. For instance, it is possible that they can move in space, thus changing their location (x_j^C, y_j^C), or they can remain in a specific location but change their pan Θ_j^P and tilt Θ_j^T parameters (Fig. 1). All cameras C_j have a sensing range R_j, and are located at a height H_j. The camera monitors a specific area which is denoted as F_j^L and represents its local (current) FoV, and is a subset of the total area that a camera is able to monitor, denoted by $\mathrm{F_j}^{\mathrm{G}}$ and $\mathrm{F_j}^{\mathrm{L}} \subseteq \mathrm{F_j}^{\mathrm{G}}$, by changing its parameters (pan, tilt, zoom, etc.). The network can consist of heterogeneous cameras that have different features such as different FoV, and motion capabilities. Finally, we assume that target associations between cameras can be established, and that ground plane information is available so that cameras can localize targets.

3.2 Target Model

Applications such as target detection and tracking are associated with a moving target that can change its position and viewpoint orientation, and thus may affect the detection performance of a camera especially as its distance from the camera increases which decreases the its pixel resolution. It is assumed that the location (x_{ji}^T, y_{ji}^T) , and distance d_{ji} of each target can be determined by the camera based on the scale size and resolution that it is detected, as well as ground plane and camera calibration information [10]. Each camera uses the above information to coordinate with other cameras regarding common target views.

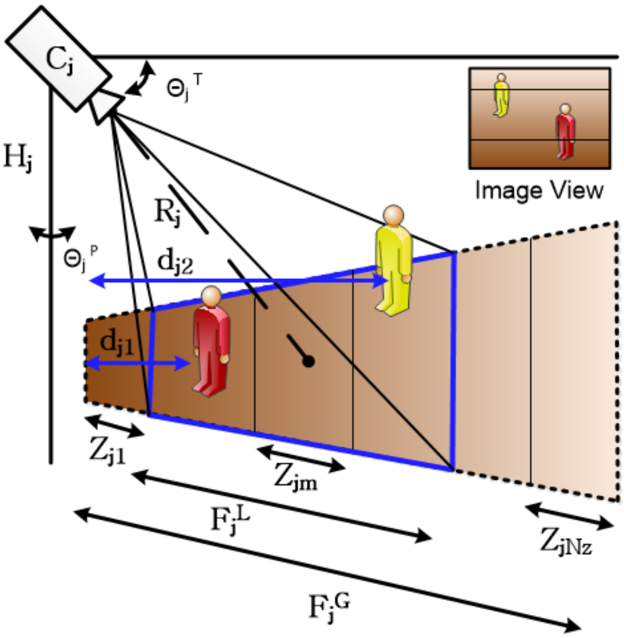

Figure 1. Active camera model and confidence/probability

3.3 Visual Sensor Uncertainty Model

The proposed uncertainty model is based on the sensing range R_j and local FoV F_j^L of each camera. Through this model, we attempt to capture how the resolution of the target in the camera image affects the probability of detection, as it will be shown in Section V. We capture this behaviour in the following way and as shown in Fig. 1:

- The global FoV $\mathrm{F_j}^{\mathrm{G}}$ of a camera C_j is segmented into m detection zones Z_{jm}, $m = 1, \dots, N_z$, where N_z is the last zone that is located further away from the camera origin.
- A camera views a subset of the m zones which belong to its local FoV $\mathrm{F_j}^{\mathrm{L}}$.
- Within each zone, there is a set of different detection probabilities. However, for simplicity we average the probabilities within the same zone and assume a uniform constant detection probability.
- The average detection probability (i.e., confidence) within a zone Z_{jm} for camera C_j is P_{jm}.
- A zone Z_{jm} has a higher detection probability if it is closer to camera location $(x_j^C, y_j^C), P_{jm} > P_{jn},\ for\ m < n.$
- A camera can establish the zone Z_{jm} that a target T_i is detected through trigonometry using the pan and tilt angles (Θ_j^P, Θ_j^T) and height H_j.
- When a target is in zone Z_{jm} it is assumed that on average is detected with rate P_{jm}.
- Different cameras can have different probability zones Z_{jm}, detection rates P_{jm} and different FoV ranges.

Algorithm 1: Zone-Based Voting Scheme

```
For $C_j \epsilon \mathcal{C}$ do
    For $T_i \epsilon \mathcal{T}$ and detected
        1. Establish $(x_{ji}^T, y_{ji}^T)$ using ground plane information
        2. Establish detection Zone $Z_{jm}$ using $(x_{ji}^T, y_{ji}^T)$
        3. Establish confidence weight $w_{jm}$ based on zone $Z_{jm}$
        4. Multicast ground plane information $(x_{ji}^T, y_{ji}^T)$ and
           weighted confidence $w_{jk}$ to other cameras
    end
end

For $C_j \epsilon \mathcal{C}$ do
If received multicast information
    If $(x_{ji}^T, y_{ji}^T)$ inside $F_j^L$
        • Aggregate information $V_j$ and establish overall decision
        • If overall decision outcome is greater than a
          voting threshold then the target is present
        end
    end
end
```

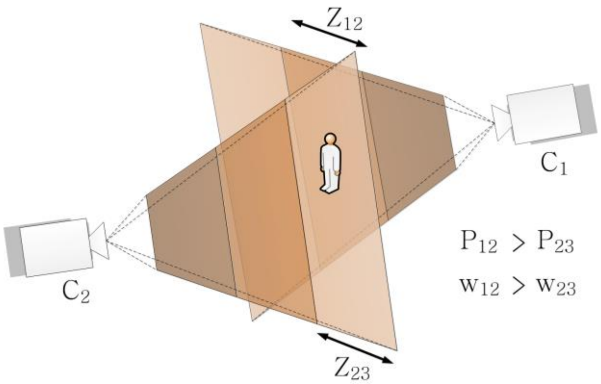

Figure 2. The vote of camera C1 will have a higher weight in the voting and combined decision making since the target is at a higher probability zone for that camera. The vote of C2 will have a lower weight since the target is at a lower probability zone for that the camera.

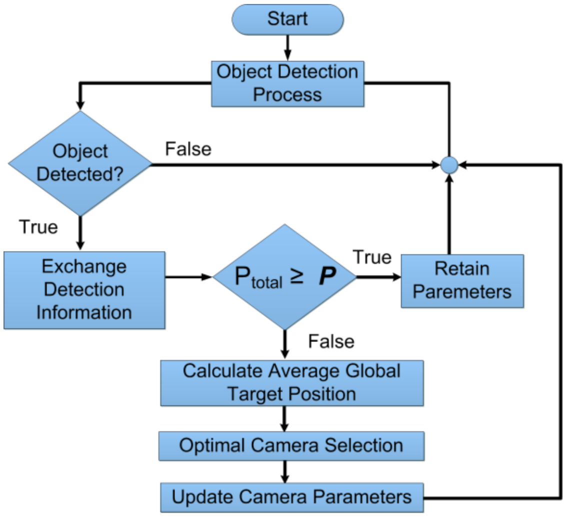

Figure 3. Algorithm for adjusting camera parameters in order to achieve the desired collective detection rate.

4. COLLABORATIVE ZONE-BASED DECISION MAKING

The presented uncertainty zone-based model, described in the previous section, can be utilized to improve the efficiency of VSNs. The detection performance degrades with the distance, change of viewpoint, and resolution of the target. As such, the probability/confidence of the camera as to how well it can detect a target depends on the position of the target. As such, the probability/confidence of a camera detection module as to how well it can detect a target depends on the position of the target. The camera detection module has a higher probability of observing targets closet to it and thus has a higher detection-rate/confidence than those for which the target is located further away. Taking this into consideration, we first propose a decision-making scheme were we use the zone Z_{jm} to determine the confidence w_{jm} with which each camera detects target $T_j$and so we accordingly weight its overall contribution to the voting process. The confidence measure w_{jm} is proportional to the detection probability P_{jm}. Along the same philosophy, we also show how to increase the overall detection probability of a network of cameras with regards to a target. We do so by selecting a number of cameras from the network that will collectively provide the highest possible detection rate. This camera selection process is directed according to the probabilities P_{jm} of each camera zone Z_{jm}.

4.1 Zone-Based Weighted Voting Scheme

We first employ the zone-based model to develop a decision-making mechanism that can be used to reach an agreement regarding the state of a target based on camera uncertainties. Through this scheme, cameras that do not detect a target due to it being in a low probability zone, will be informed by other cameras which detect the target at a higher probability zone and remain aware of the targets state. It is based on a voting scheme (outlined in *Algorithm 1*) for collaborative decision making where each camera camera C_j maintains a voting vector V_j for T_j to store the decision of neighbouring cameras. First, a target T_j enters the F_j^L of a camera C_j and is positioned in a zone Z_{jm}. Then the target can be detected with a probability P_{jm} and localized by each camera, using ground plane information, to a position (x_{ji}^T, y_{ji}^T). This information is transmitted to other cameras which also estimate its position which if it is outside the F_j^L then P_{jm} is zero and they do not take part in the voting process. Otherwise, the camera outcome is weighted with a zone weight w_{jm}, and is multicast to the other cameras. An example of this is illustrated in Fig. 2. Once this information is received by all cameras, they update decision vector V_j, aggregate it, and threshold the result to reach to an agreement regarding the state of a target.

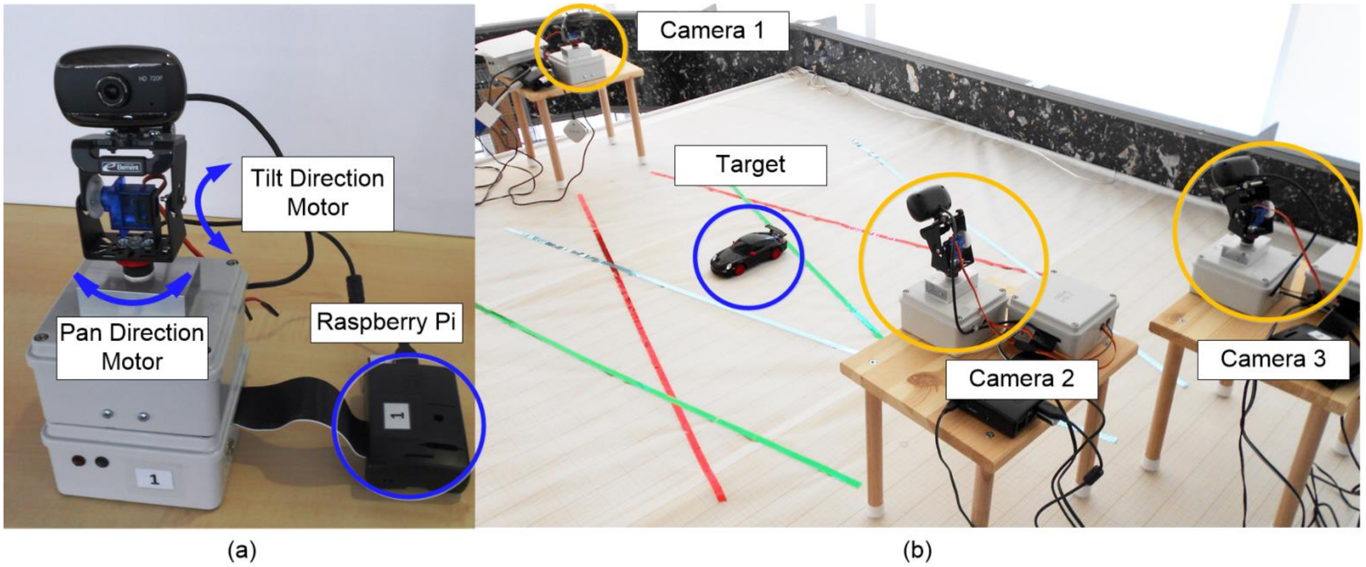

Figure 4. Experimental Setup: (a) The developed Raspberry-Pi pan-tilt smart camera station. (b) The network of cameras collaboratively monitoring a target. Notice that the cameras (e.g. camera 3) can have different initial orientations.

4.2 Zone-Based Configuration Mechanism for Active Vision

In this section, we show another application of the proposed model where cameras can request the assistance of other cameras in the network, whenever a target is in a lower confidence zone and an acceptable decision confidence is not achieved. In this event, one or more of the assisting cameras will need to adjust their state in order to meet the confidence requirements. The overall process is outlined in Fig. 3, and is initiated once one of the cameras detects an object. First, the cameras exchange information regarding the object they detect such as position coordinates and detected zone. Each zone is characterized by a detection confidence, which is the probability that a target can be detected. This detection confidence can help determine whether other cameras also need to monitor the target. If one or more cameras detect an object with a collective confidence P_{total} that is less than a predetermined threshold P (i.e., the object will not be detected with the required probability), then if other cameras are available one of them is selected to participate in the detection of that object in order to increase the overall confidence. The camera that will be selected can be chosen in different ways depending on the objective. For example, the objective can be to select the camera which will require the least amount of movement in order to conserve energy, or it can be to select the one closer to the target which will add the highest amount to the collective confidence. Regardless of the criterion, the selected camera will then update its parameters to move to the location where it will monitor the target. If one camera is not sufficient to achieve the desired confidence then the process is repeated again and so more cameras are added until it is met. Of course, an acceptable solution may not always be feasible due to the initial camera placement, in which case, the cameras will reconfigure to achieve the highest possible collective detection probability.

5. EXPERIMENTAL SETUP & EVALUATION RESULTS

To evaluate the proposed model and decision-making process we have developed a network of smart cameras based on the Raspberry Pi single-board computer [15]. Each Raspberry Pi is connected with a webcam that is mounted on a motorized two degrees-of-freedom (DoF) pan-tilt stage, as shown in Fig. 4-a. The two angular positions are controlled independently using a corresponding servo motor and they are equipped with potentiometer-type position sensors. The sensory feedback information allows accurate angular positioning of the pan-tilt system. The servo motors are controlled by the Raspberry Pi and the control electronics using a pulse width modulation (PWM) approach. Communication between the camera stations is realized via a dedicated local Wi-Fi network. Each camera station is also fitted with programmable LEDs that indicate the status of the system (e.g., object detected). A grid field ($1{,}5 \times 1{,}5\, m^2$) was used over the surface where the cameras were positioned, that provided a global coordinate system and facilitated registration between the reference frames corresponding to each camera station. The cameras were also able to calculate an estimate of the targets position in a global reference system using trigonometry and the current angle configurations. The target objects were remote controlled cars. For this reason, we trained an image classifier capable of detecting cars using the Cascade Object Detection Algorithm with Local binary Pattern (LBP) features based on the seminal work by Viola and Jones [16] which is available in the OpenCV computer vision library [17]. The training set was constructed using the database from [18] and was enhanced with additional sample images. The experiments were conducted in non-controlled environments with ambient light.

5.1 Zone-Based Confidence Model Evaluation

We first evaluated the validity of the proposed uncertainty model using the developed Raspberry-Pi smart camera station. The station was configured with different orientations (looking straight and at different tilt angles) and for each one the detection rates of the target object were measured and averaged for multiple runs for 100 consecutive frames, at different positions covering the whole field. The results for these experiments are shown in Fig. 5, where we first show the effective FoV area that the camera can monitor out of the $25cm \times 16cm$ area in front of the camera. Notice how the effective FoV is different for angled and straight looking cameras. Also Fig. 5 illustrates the detection probabilities within the effective FoV of a camera. Notice that as the distance of the

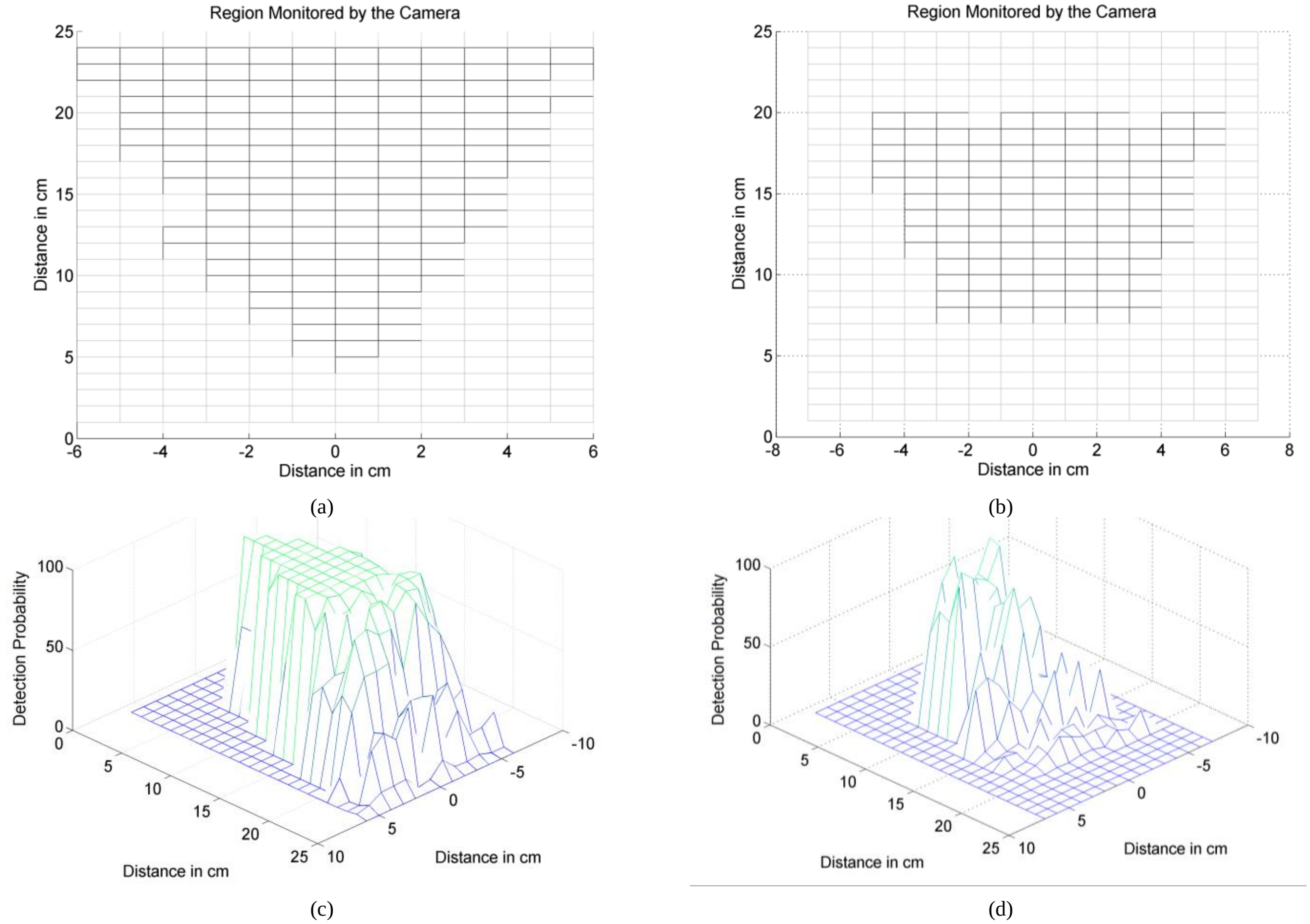

Figure 5. (a) Field Area Monitored by straight looking cameras. (b) Field Area Monitored by angled camera. (c) Detection probability degradation for straight looking cameras. (d) Detection probability degradation for angled camera

target increases the detection rate deteriorates. This is because as the object resolution decreases, and is represented with less pixels, slight variations in a few pixels can cause the detection module to produce the wrong outcome. Of course, the dimensions and specific detection rates depend on the camera resolution as well as the detection algorithm itself. We have used the state-the-art Cascade detection algorithm found in the OpenCV computer vision library [17] which is a typical and widely used example of a detection algorithm. Hence, the general trend is expected to remain and the model to be applicable for different camera configurations, with only the actual probability values changing which would require some additional experiments in order to be determined. Using the model, we can then extract the probability zones that can be used for camera coordination purposes.

5.2 Application to Active Vision

The extracted values and model have been used in an active vision experiment. For this experiment, we used three smart camera stations that communicated wirelessly in order to exchange information and coordinate their actions. Information exchanged between the stations included a notification with the camera's ID each time an object was detected, the object's coordinates (derived from its position in the image and the joint rotations of the pan-tilt stage) as well as the detection probability for the object (corresponding to the spatial zone in which it was detected). The target was placed in various positions within the fields and the cameras responded by accordingly adjusting their configurations in order to meet the necessary overall detection probability. Cameras were placed at the same height at arbitrary initial orientations. At every station, the control computer executed a program implementing the abovementioned algorithm in order to reconfigure according to the previously described process (Fig. 3).

Each time a target is identified by at least one camera the network reacts appropriately. When an object is detected with a satisfactory probability then the camera configurations are maintained. Otherwise, in the case where the object was positioned and detected in a camera zone with a non-satisfactory detection probability, the cameras initiate the new camera selection process. After processing the available information, the minimum required number of cameras is selectively employed to ensure the desired detection probability using the proposed zone model. The selected camera(s) were able to determine their new pan and tilt angles by using trigonometry and the calculated distance of the target.

Based on the analysis in the previous section we employ a 3-zone model and the assigned detection probabilities were 90% for the proximal zone, 55% for the intermediate zone and 25% for the distant one. These parameters were the same for all cameras. In Table I three representative test cases are reported where the target position was varied. The table also indicates the zone probability values of the cameras when the target is present in their F_j^L. Hence, in each case the system had to employ a different

Table 1. Theoretical and Experimental Detection Rates (%)

	Theoretical (%)				Experimental (%)			
	Camera 1	**Camera 2**	**Camera 3**	**Total**	**Camera 1**	**Camera 2**	**Camera 3**	**Total**
Case 1	90	-	-	90	99	-	-	99
Case 2	55	55	-	79	51	44	-	71
Case 3	55	25	25	74	68	17	32	79

number of cameras in order to provide the required detection probability. The required detection probability in the experiments was set to 80%, however, this threshold can change depending on the targeted application. The theoretical total detection rate achieved by the cameras can be calculated using probability theory, given that the detection events generated by the cameras are independent. The experimental combined detection rates were measured by combining the detection results of the cameras for 100 consecutive frames and averaging for multiple runs, after the new configuration was set. In the first case, camera 1 was enough to monitor the target with sufficient rate. In the second case camera 2 detects the target at zone 2 with a lower detection rate than required. Hence, Camera 1 (which has the minimum distance) is selected to add the remaining rate. Camera 3 did not view the target so the maximum possible rate was reached. Finally, in the last case camera 1 detects the target and both 2 & 3 were selected to contribute with their corresponding probability based on the target position in the respective zones. Since the model zone probabilities are the average of each region there is the possibility that we may go over or under the theoretical value, as in case 1 and 2. In addition, some detection events may overlap between the cameras and hence the theoretical maximum may not be achieved, as in case 3. Overall, the experiments verified the validity of the proposed zone-based model and how it can be used to direct decision making and reconfiguration in VSNs.

6. CONCLUSIONS

This paper presented research towards the realization of a camera uncertainty model that can be used to improve the robustness and reliability of VSNs for various applications. We have demonstrated the validity of this model and how it can be utilized in order to collaboratively and dynamically configure a network of smart cameras to maintain a predefined detection probability. The model presented herein can serve as the basis for future work and can be further developed to capture and characterise the behaviour of smart cameras in VSNs. The effort going forward will be on enhancing the model by compensating for occlusions and handling false detections.

7. ACKNOWLEDGMENTS

This work was supported in part by the ERC Advanced Grant "Fault-Adaptive", ERC grant agreement nº 291508.

8. REFERENCES

[1] Aghajan H. and Cavallaro.A. 2009. *Multi-Camera Networks: Principles and Applications*. Academic Press.

[2] Qureshi, F. and Terzopoulos D. 2007. Camera Sensor Networks in Virtual Reality. *First ACM/IEEE International Conference on Distributed Smart Cameras*. (25-28 Sept. 2007). ICDSC 2007. 87-94.

[3] Rinner B. and Wolf B. W. 2008. An Introduction to Distributed Smart Cameras. *Proceedings of the IEEE*. 96, 10 (Oct. 2008), 1565-1575.

[4] Mahmut Karakaya and Hairong Qi. 2014. Collaborative localization in visual sensor networks. *ACM Trans. Sen. Netw.* 10, 2, Article 18 (January 2014), 24 pages. DOI=http://doi.acm.org/10.1145/2529999

[5] Rowe A G., Goode A., Goel D., Nourbakhsh I. 2007. *CMUcam3: An Open Programmable Embedded Vision Sensor* Technical Report, Carnegie Mellon University.

[6] Sunderrajan S., Manjunath B.S. 2013. Multiple view discriminative appearance modeling with IMCMC for distributed tracking. *International Conference on Distributed Smart* Cameras (29 Oct.-1 Nov. 2013). ICDSC 2013. 1-7.

[7] Micheloni C., Rinner B., Foresti G.L. 2010. Video Analysis in Pan-Tilt-Zoom Camera Networks. *IEEE Signal Processing Magazine*. 27, 5 (Sept. 2010). 78-90.

[8] SanMiguel J.C., Micheloni C., Shoop K., Foresti G.L., Cavallaro A. 2014. Self-Reconfigurable Camera Sensor Networks. *Computer*. 47, 5 (May 2014). 67-73.

[9] Dinh T., Yu Q., Medioni G. 2009. Real time tracking using an active pan-tilt-zoom network camera. In *Proceedings of the 2009 IEEE/RSJ international conference on Intelligent robots and systems* (IROS'09). 3786-3793.

[10] Lim S., Elgammal A., Davis L. S. 2003. Image-based Pan-tilt Camera Control in a Multi-Camera Surveillance Environment. *IEEE International Conference on Multimedia and Expo*. (July 2003). 6-9.

[11] Biswas A., Guha P., Mukerjee A., Venkatesh K.S. 2006. Intrusion Detection and Tracking with Pan-Tilt Cameras," *IET International Conference on Visual Information Engineering*. (Sept. 2006). 565-571.

[12] Kamal A.T., Farrell J.A., Roy-Chowdhury A.K. 2013. Information Consensus for Distributed Multi-target Tracking. *IEEE Conf. on Computer Vision and Pattern Recognition*. (23-28 June 2013). CVPR 2013. 2403-2410.

[13] Kamal AT., Farrell J.A., Roy-Chowdhury A.K. 2013. Information Weighted Consensus Filters and Their Application in Distributed Camera Networks. *IEEE Transactions on Automatic Control*. 58, 12 (Dec. 2013), 3112-3125.

[14] Morye A.A., Ding C., Roy-Chowdhury A.K., Farrell J.A. Distributed Constrained Optimization for Bayesian Opportunistic Visual Sensing. *IEEE Transactions on Control Systems Technology*, 226 (Nov. 2013). 2302-2318.

[15] Raspberry Pi. Available Online: http://www.raspberrypi.org/

[16] Viola P., Jones M. J. 2004. Robust Real-Time Face Detection. *Int. J. Comput. Vision* 57, 2 (May 2004), 137-154. DOI=http://dx.doi.org/10.1023/B:VISI.0000013087.49260.fb

[17] Bradski G.. 2000. The OpenCV Library. Dr. Dobb's Journal of Software Tools, Available Online: http://opencv.org/

[18] Agarwal S., Awan A., Roth. D. 2004. Learning to Detect Objects in Images via a Sparse, Part-Based Representation. *IEEE Trans. Pattern Anal. Mach. Intell.* 26, 11 (Nov. 2004), 1475-1490. DOI=http://dx.doi.org/10.1109/TPAMI.2004.108