

# Wireless Multimedia Sensor Network deployment for disparity map calculation

Anthony Tannoury<sup>a,b</sup>, Rony Darazi<sup>a</sup>, Abdallah Makhoul<sup>b</sup>, and Christophe Guyeux<sup>b</sup>

<sup>a</sup>TICKET Lab, Antonine University, Hadat-Baabda, Lebanon.

<sup>b</sup>Univ. Bourgogne Franche-Comté, FEMTO-ST Institute/CNRS, Belfort, France

Emails: anthony.tannoury@ua.edu.lb, rony.darazi@ua.edu.lb, abdallah.makhoul@univ-fcomte.fr, christophe.guyeux@univ-fcomte.fr

**Abstract**—The use of Wireless Multimedia Sensor Networks (WMSNs) has been increased in the recent decade in several fields. One of the main advantages of WMSNs is activity monitoring, where 2D images and videos have been used traditionally for surveillance and detection so far. However, a lot of interest is paid to depth information that is not normally available while using 2D images. In our previous work, the use of disparity information in WMSNs has been studied. In this paper, our contribution consists of evaluating the Sensor deployment for efficient disparity calculation for indoor and outdoor purposes. The simulation results show that while considering images from adjacent sensors, the disparity map is successfully calculated. The quality of the disparity is evaluated using structural similarity metric between one reference image captured by one sensor and the calculated disparity map.

**Index Terms**—Wireless Multimedia Sensor Networks, Disparity Map, Stereo Vision, 3D Scene Reconstruction, Surveillance, Event Detection.

## I. INTRODUCTION

Scene surveillance is a promising research subject in Internet of Things (IoTs). Traditional Wireless Sensor Networks (WSNs) can sense scalar data like humidity, speed, etc [1]. The added value(s) in WMSNs is the multimedia data that can be sensed and delivered such as image, video, and audio. The availability of small-scale and low-powers Complementary Metal Oxide Semiconductor (CMOS) camera sensors helps in developing new WMSNs applications in multiple fields like military, e-health, etc [2]. Requirements in terms of quality of service, energy consumption and computations time differ from one application to another. In WMSNs surveillance systems, the main challenge is to develop an efficient monitoring network that covers the maximum range of target objects with the lowest computation time and energy consumption.

Calculating the depth value in 3D space is an essential indicator before detecting a change in the monitored scene and then triggering an alert or 3D reconstructing the scene. Actual stereo vision systems and 3D reconstruction solutions lies on professional and complex hardware like laser scanners or advanced cameras [3]. In [4], even transparent objects could be recognized using Intel Realsense R200 long range stereoscopic depth sensor. Guizi et al. [5] proposed a 3D algorithm, via tri-dimensional depth estimation, for real time people detection and tracking in indoor environments. However, video sensors in WMSNs are limited in power resources because they are

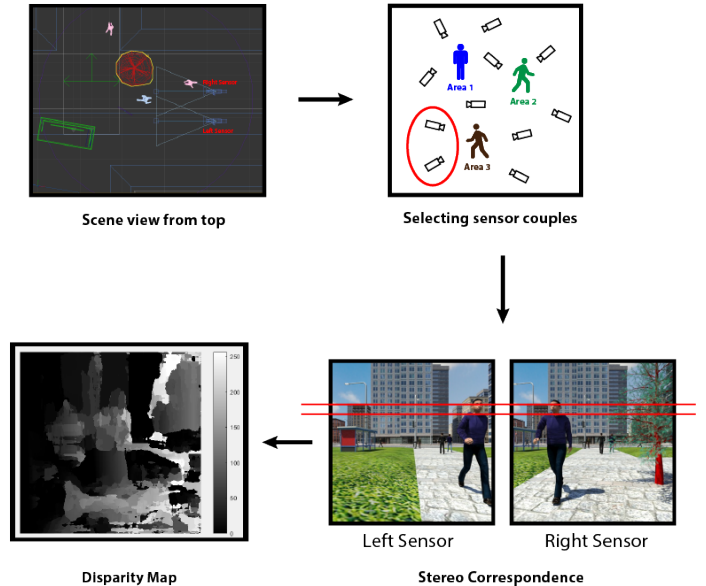


Fig. 1. Proposed model

powered by batteries. Our main objective is to calculate the 3D depth while saving power and then delivering valuable multimedia data to the sink(s), sufficient for event/changes detection or 3D Scene Reconstruction [6].

For the above-mentioned reasons, we introduced the disparity map computation in the context of WMSNs based surveillance [7], [8]. The depth information is saved in the disparity maps calculated on each couple of sensors, marked as left and right cameras. The Fig. 1 outlines our system. As can be seen, the double-view images are processed locally via the couple of low cost sensor nodes. This will reduce transmission rate, which is the direct and important cause for altering the network lifetime. Our system will not deliver high size images but only gray-scale disparity maps, on demand or triggered by a change or intrusion in the monitored scene/object.

This remainder of the paper is organized as follows. Section II explains the context of surveillance using WMSNs. Section III introduces the disparity map usage for WMSNs surveillance systems, respecting WMSNs specifications, challenges and limitations. In Section IV, one low cost disparity map calculation method is experimented, particularly the Sum

of Absolute Differences (SAD). The experimentation is done with reference to our previous work [7], where SAD achieved the best results in terms of complexity, computation cost, and speed. Two different scenes, indoor and outdoor, help us to determine the system configuration and sensor nodes deployment with different angles/views and distances. This leads to extract good and well structured disparity maps, holding a high percentage of scene details. Finally we conclude our paper in section V.

## II. BACKGROUND

WMSNs are used in different domains and applications, popularly implemented for surveillance [9]. The development of new surveillance systems is increasing and it motivates both researchers and industrial experts. They try to reach higher level of safety in different contexts such as national security, traffic monitoring, habitat monitoring, e-health, etc [10]. Traditional single view monitoring systems recommend a 24/7 human intervention in order to detect or track an intrusion/changes. Adding to this, the high cost and complexity of implementation to configure a well designed hardware based surveillance system must be taken into account. Using WMSNs is very advantageous because of sensors mobility and multiple view cameras. A single camera cannot observe the complete area of interest. Multiple image sensors sources give the application the opportunity to adaptively process and extract data on demand. So WMSNs will cover more areas and bigger ranges for automated real-time monitoring. WMSNs nodes are able to capture, store, process and aggregate multimedia data transmitted from different sensors in the network [11].

We are focusing on developing a WMSNs system to monitor a scene from different views, using low cost cameras, that can cover and reconstruct the maximum possible area. We will show in Section III how we can profit from different sensor couples to calculate disparity maps holding depth values and essential for 3d scene monitoring or reconstruction. In our previous research work [12], we investigated all WMSNs applications and constraints from disparity map calculation perspective, where we studied all disparity map methods and chosen possible algorithms applicable in our context.

Energy consumption is the main constraint while developing new WMSN systems independently from the applications like monitoring, objects tracking, traffic avoidance, etc. Transmission consumes the big portion of the power, as experimented in [13]. So decreasing the transmitted multimedia data size and distance between sensors leads to increasing network lifetime. Our distributed WMNSs will calculate disparity maps on each sensor couples and deliver to the sink low size gray-scale disparity maps.

Another crucial constraint in WMSNs is the coverage problem, where random objects like trees or cars can cause occlusion with the monitored target. Also, two parallel sensors can be overlapped together. Figure 2 shows an example of occlusion and overlapping. Costa et al. [14] surveyed this specific issue. The next sections III and IV will take into

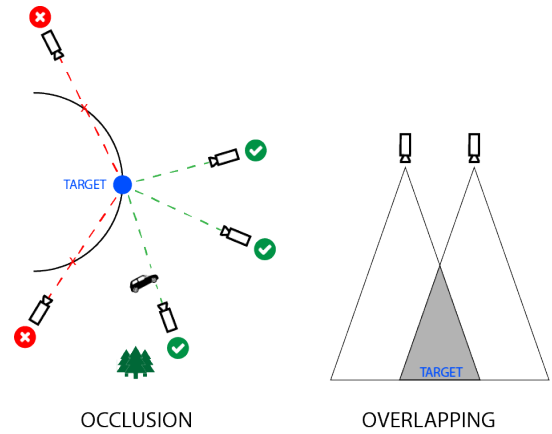


Fig. 2. Occlusion and overlapping problems in video sensor network coverage

consideration all conditions to ensure a good camera sensors placement for better disparity map calculation. This plays an important role by putting sensors in sleep mode or reactivating them based on the area of interest, thus extending the network lifetime.

Consequently, even with these constraints, WMSNs are very advantageous for surveillance purpose including:

- 1) Enlarging the view: two different camera nodes can monitor the same target but from different distances, one closer from the second;
- 2) Large field of view: different camera with multiple Field Of Views (FOVs) can be used;
- 3) Different views: multiple cameras, at the same distance from the target, can point to it from different angles.

In the next section, referencing to our previous contributions [12], [7], we will explain disparity map calculation constraints respecting to WMSNs prementioned limitations. This ensure an efficient usage of WMSNs for disparity map calculation.

## III. DISPARITY MAPS IN WIRELESS MULTIMEDIA SENSOR NETWORKS

The main idea of using disparity maps via WMSNs came from before-mentioned network limitations and the low size of these black and white maps. Calculating the depth value(s) is the essential key to trigger any alert after scene change(s). Sensor nodes can enter sleep mode, so in case an event occurs, sensors can wake up each others to recalculate again new disparities. So worst case, one sensor will be active and the second one sleeps thus decreasing 50 % of the processing and transmission on distributed pairs level. In our last article [12], we surveyed all disparity map calculation methods and regrouped them in terms of:

- Computational demand: clock frequency (in MHz/GHz), frame rate (FPS), image quality (in pixels).
- Local or Global method: global methods bring off better results than local but are computationally expensive.
- Disparity Range: some methods reach deeper disparity ranges.

- Occlusion, discontinuities and uniform textures or texture-less areas handling.
- Hardware implementation: requires a graphical processing unit GPU, central processing unit CPU or field-programmable gate array FPGA.

Our proposed compromise operates as follows: calculate efficient and accurate disparity maps on a couple of sensors, with limited computational and energy resources. We chosen to use local methods with the minimum clock rate and image quality. Thus we compared and experimented two acceptable methods: Sum of Absolute Differences (SAD) and Sum of Squared Differences (SSD) [15] [7]. Processing time increased linearly with images resolutions and number of couples (pairs of sensors).

Let us recall that the SAD algorithm calculates the absolute difference between the intensity of each pixel in the reference block and that of the corresponding pixel in the target block:

$$SAD(x, y, d) = \sum_{(x,y) \in w} |I_l(x, y) - I_r(x - d, y)| \quad (1)$$

- $(x, y, d)$  represent the disparity map coordinates;
- $(x, y)$  are the coordinates of the pixel of interest;
- $d$  is the disparity value;
- $I_l$  is the left reference image;
- $I_r$  is the right target candidate image.

It makes the sum of differences over  $w$ , where  $w$  is the aggregated support window. SAD over-performed SSD with lower computational complexity and processing time, so it is convenient for real-time applications.

Studying, experimenting, and resolving computational requirements while choosing SAD as ideal disparity map calculation method for WMSNs monitoring is principal. But, as stereo matching has its own constraints, this paper outlines all stereo vision requirements for better sensor deployment. At this stage, we can wonder: what are the restrictions to compute good quality disparity maps, and monitor an outdoor or indoor scene from different views and perspectives?

For all researchers working in this field, designing and developing accurate stereo vision systems, and improving the quality of 3D projections, stereo baseline is preliminary. This is the distance between sensor 1 and sensor 2. Wutthigrai Boonsuk [16] implemented a stereo vision system for industrial robotics and concluded the following: shorter baseline distances performs better at shorter distances to the target, but at longer distances, wide baselines are better. Si et al. [17] used camera sensor networks to detect intrusions passing through a barrier area. They implemented multiple parallel cameras in order to guarantee a maximum coverage for the monitored barrier. Increasing the number of cameras lets them ensure that a minimum of two sensors will capture the intruder.

Kim et al. [18] proposed a 3d modeling method using high resolution spherical images. Their 3D reconstruction was based on stereo image pairs using vertical displacement between camera views. The epipolar geometry depends on the

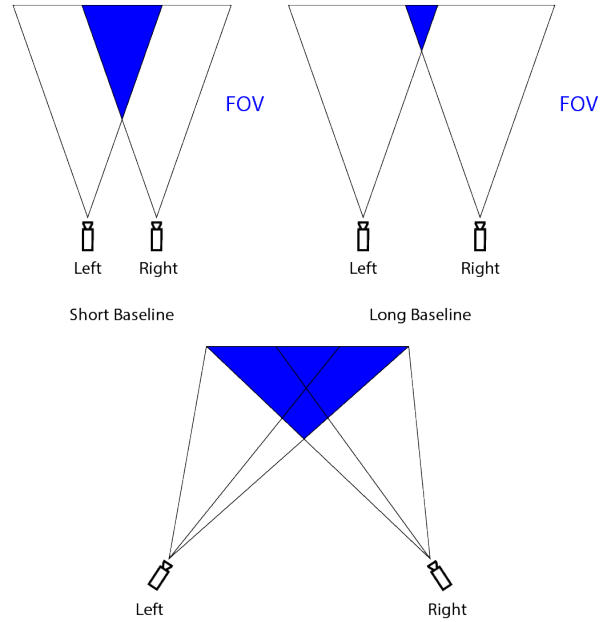


Fig. 3. Stereo vision system baseline and orientation

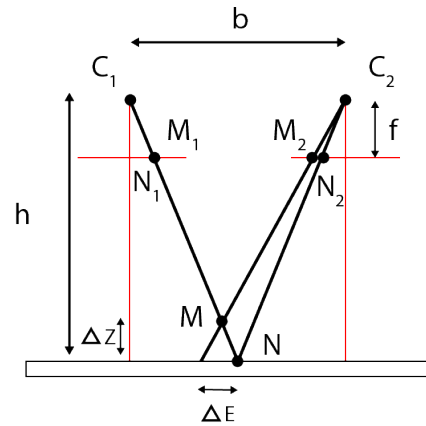


Fig. 4. Stereo Coefficient principle

cameras extrinsic and intrinsic parameters. Extrinsic parameters are the camera rotation and translation in the environment. Intrinsic parameters are the internal properties of the camera itself like focal length, pixel width, and pixel height.

In our context, the added value is computing the disparity map between two camera nodes and tracking the change(s) in depth or reconstruct the intruder in the scene.

Figure 3 illustrates the effect of increasing parallel cameras baseline or rotating them. The disparity value is between 0 and 1 and it increases with the sensing direction. So a high disparity value indicates a high difference between the couple of images: they are less correlated. Otherwise, two perpendicular cameras, weakly correlated, leads to a disparity of 1 [19].

Figure 4 shows how depth changes with baseline  $b$  and distance to target  $h$ .  $C_1$  and  $C_2$  are the centers of the sensors.  $M_1$  and  $N_1$  are the projections of the points  $M$  and  $N$  in the



Fig. 5. Indoor scene: elderly man walking in his private bedroom



Fig. 6. Outdoor scene: random populations walking in city streets

first image.  $M_2$  and  $N_2$  in the second image.  $\Delta M$  and  $\Delta N$  are the shifts between these positions.  $\Delta M - \Delta N$  is proportional to the disparity  $\Delta E$  and  $\Delta E$  itself is approximately proportional to the depth difference  $\Delta Z$ . A larger coefficient  $b/h$  leads to smaller error in depth. However, a high coefficient reflects more changes between the images, hence more difficult matching process. For example, metropolitan images where there is a big number of buildings creating occlusions and changing fast with observation angle. Thus, smaller angles between views is recommended for accurate disparity maps.

So the choice of the coefficient  $b/h$  depends on the monitored scene scenario compromise. Delon et al. [20] studied deeply stereo visions with small baselines and resulted that an angle of 53%, typically  $b/h = 1$ , is ideal for indoor. Objects far from sensors lenses shift less, and objects nearer shift more. In traditional WMSNs coverage, it is advantageous to increase the angles and distances between the nodes to avoid duplication of data. Two near sensors may monitor the same object. But in our context, it is different because we tend to increase intersection between cameras, since we risk to search for a match in the second sensor that is not shown in the first reference one. The next section shows experimental results for two different outdoor and indoor scenes, where disparity calculation is made with different baselines, distances and angles. This ensure and improve the sensors deployment strategy for better disparity maps quality.

#### IV. EXPERIMENTATION RESULTS

We created our own data-set using Autodesk 3ds Max 2018, where we modelled two realistic outdoor and indoor scenes. The indoor scene represents an elderly man walking in the bedroom, as shown in Figure 5. Similarly, Figure 6 displays

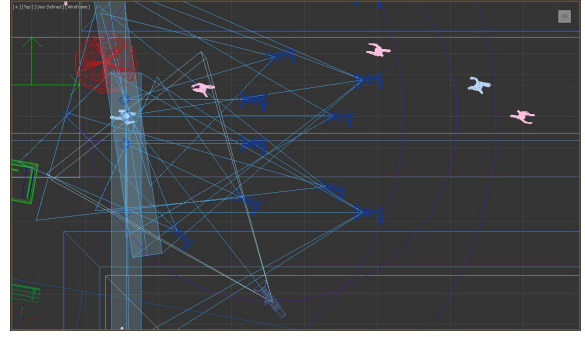


Fig. 7. Camera positions from top

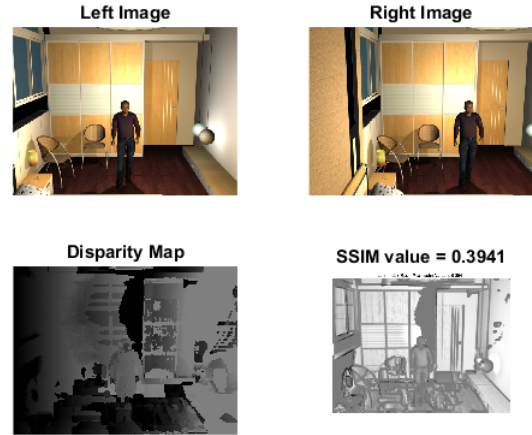


Fig. 8. Applying SSIM on a computed disparity map of two different views

the outdoor scene of an urban city where multiple persons are promenading in the street. This represents two different WMSNs E-Health and habitat monitoring applications.

The simulated scenes have different textures, lighting conditions, colors, reflections, and occlusions. So it keeps natural and random environmental conditions as real as possible. Cameras are placed with predefined field of views (FOV) of 45 and 60 degrees. Our experimentation takes into consideration multiple baseline distances and distance to the target (the man) in meter(s), and angles of observation in degree(s). Figure 7 shows an example of camera sensors placement to cover the monitored scene.

Firstly, for each camera couple, we rendered the images and exported them as Joint Photographic Experts Group (JPEG). The rendered captured images are low in resolution of 640x480 pixels (indoor) and 480x270 pixels (outdoor). This is because resolutions affects directly the disparity map processing time, as experimented in our last publication [7]. Secondly, Sum of Absolute Differences (SAD) uses them as parameters to compute disparity maps via MATLAB 2017. Finally, we made a quality testing on the computed disparity maps to conclude the best sensors placements (distance and orientation). To experiment the disparity maps quality, tradi-



TABLE I  
SSIM VALUES FOR DIFFERENT BASELINES AT 1.5 M FROM TARGET  
(INDOOR)

Perfectly aligned sensors		
Baseline b (meters)	Distance to target h (meters)	SSIM
0.1	1.5	0.1363
0.2		0.1596
0.3		0.1947
0.4		0.2350
0.5		0.2529
0.6		0.2537
0.7		0.2395
0.8		0.2364
0.9		0.2637

tional computer vision researchers use a ground truth disparity map as reference. This best quality reference is computed with a laser scanner or from middlebury data-set [21]. This is not feasible in our context since we created our own virtual scenario, so no ground truth.

The Structural Similarity (SSIM) is used to solve this issue and define a disparity map quality measurement methodology. SSIM measures the similarity between two images, thus making it convenient to compare the saved structure between the computed disparity and one of the left or right images pair. To prove the concept, we applied SSIM on different disparity maps going from bad to very high quality, and we monitored how SSIM changed linearly from 0 to 1.

Figure 8 illustrates a high SSIM of 0.3941 of a good disparity map between two sensors having a baseline of 0.6 meters and 3.0 meters from the target (walking man). Table I, Table II, and Table III show different disparity map qualities for multiple baselines, angles and distances for indoor and outdoor scenes. Table I shows how SSIM increases with the baseline distance. So, for indoor scenes, it is recommended to use wide baselines thus ensuring a small FOV, small depth error but more occlusions risks. For small baselines, the stereo correspondence is simpler as images are similar with few occluded regions, but it risks a large depth uncertainty. Based on Table II results, increasing the angle between the sensors gives us more opportunity to have good disparity maps for long distance to target. For short distances, changes in the angle results a big difference in the views. Table III insured that for outdoor scenes, monitoring a long ranged target can be done with higher angle between sensors. Yet, the ideal baseline for indoor monitoring is 0.9m. The best angle for rotated sensors is 30 degree with a baseline of 3.0m. For outdoor, putting sensors with 10 degrees rotation and 6m to target calculates good disparity maps.

## V. CONCLUSION AND FUTURE WORK

Calculating disparity maps on distributed sensors is a WMSNs added value for traditional multimedia data by saving and delivering depth information. This is preliminary and essential for events/changes detection or 3D scene reconstruction. Our approach respects all WMSNs limitations, specially power consumption and quality of service. This research outlines an

TABLE II  
SSIM VALUES FOR DIFFERENT ANGLES AT DIFFERENT DISTANCES FROM TARGET (INDOOR)

Rotated sensors		
Angle between sensors (degree)	Distance to target h	SSIM
30	0.9	0.2539
	1.5	0.3377
	3.0	0.3649
90	0.9	0.1754
	1.5	0.2237
	3.0	0.2640

TABLE III  
SSIM VALUES FOR DIFFERENT ANGLES (OUTDOOR)

Angle between camera1 and camera2 (degree)	Distance = 4 m	Distance = 6 m	Distance = 8 m
10	0.1597	0.1602	0.1453
20	0.1318	0.1346	0.1282
30	0.1387	0.1445	0.1367
40	0.1371	0.1356	0.1344
50	0.1375	0.1294	0.1431
60	0.1395	0.1318	0.1506
70	0.1310	0.1386	0.1502
80	0.1235	0.1354	0.1533
90	0.1293	0.1354	0.1575

experimental contribution to ensure good sensors deployment for ideal disparity map quality.

For future work, we intend to explore and define how disparity maps will be transferred from couple to another, from nodes to aggregator(s). In this way, we will reach an efficient WMSNs based surveillance system that covers a high range of targets for long lifetime. The 3D reconstructed information(s) can help different applications in many domains like elderly people monitoring, military tracking, habitat monitoring, etc.

## ACKNOWLEDGMENT

*This work is partially funded by the Labex ACTION program (contract ANR-11-LABX-01-01), the France-Suisse Interreg RESponse project, the National Council for Scientific Research in Lebanon CNRS-L and the Hubert Curien CEDRE programme.*

## REFERENCES

- [1] N. Brindha and S. Vanitha, "The data aggregation approach to enhance the lifetime of wireless sensor network," in *2017 International Conference on Innovations in Green Energy and Healthcare Technologies (IGEHT)*, March 2017, pp. 1–5.
- [2] V. Ukani, P. Thakkar, and V. Parikh, *Routing Protocols for Wireless Multimedia Sensor Networks: Challenges and Research Issues*. Cham: Springer International Publishing, 2018, pp. 157–164.
- [3] O. Ripolles, J. E. Simó, G. Benet, and R. Vivó, "Smart video sensors for 3d scene reconstruction of large infrastructures," *Multimedia Tools and Applications*, vol. 73, no. 2, pp. 977–993, Nov 2014. [Online]. Available: <https://doi.org/10.1007/s11042-012-1184-z>
- [4] Y. Yun, D. Seo, and D. Kim, "Recognition of transparent objects using 3d depth camera," in *2017 14th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI)*, June 2017, pp. 882–883.

- [5] F. da Silva Guizi and C. S. Kurashima, "Real-time people detection and tracking using 3d depth estimation," in *2016 IEEE International Symposium on Consumer Electronics (ISCE)*, Sept 2016, pp. 39–40.
- [6] J. M. Bahi, C. Guyeux, A. Makhoul, and C. Pham, "Low cost monitoring and intruders detection using wireless video sensor networks," *International Journal of Distributed Sensor Networks*, vol. 2012, 2012.
- [7] A. Tannoury, R. Darazi, C. Guyeux, and A. Makhoul, "Efficient and accurate monitoring of the depth information in a wireless multimedia sensor network based surveillance," in *To appear in the proceedings of International Conference on Sensors, Networks, Smart and Emerging Technologies SENSET17*, 2017.
- [8] J. Bahi, C. Guyeux, A. Makhoul, and C. Pham, "Secure scheduling of wireless video sensor nodes for surveillance applications," in *3rd Int. ICST Conference on Ad Hoc Networks*, vol. 89, 2011.
- [9] S. A. Sert, A. Yazc, and A. Coar, "Data fusion and processing in wireless multimedia sensor networks: An analysis for surveillance applications," in *2014 22nd Signal Processing and Communications Applications Conference (SIU)*, April 2014, pp. 421–424.
- [10] O. Javed and M. Shah, *Automated Multi-Camera Surveillance: Algorithms and Practice*, 1st ed. Springer Publishing Company, Incorporated, 2008.
- [11] L.-m. Ang, K. P. Seng, L. W. Chew, L. S. Yeong, and W. C. Chia, *Wireless Multimedia Sensor Networks on Reconfigurable Hardware: Information Reduction Techniques*, 1st ed. Springer Publishing Company, Incorporated, 2016.
- [12] A. Tannoury, R. Darazi, A. Makhoul, and C. Guyeux, "Introducing disparity map for 3d scene reconstruction in wireless multimedia sensor networks," May 2017, technical report and submitted article.
- [13] Q. Zhao and Y. Nakamoto, "Estimating the battery consumption of data processing in a wireless sensor node," in *2013 First International Symposium on Computing and Networking*, Dec 2013, pp. 484–486.
- [14] D. G. Costa and L. A. Guedes, "The coverage problem in video-based wireless sensor networks: A survey," *Sensors*, vol. 10, no. 9, pp. 8215–8247, 2010. [Online]. Available: <http://www.mdpi.com/1424-8220/10/9/8215>
- [15] A. Tannoury, R. Darazi, C. Guyeux, and A. Makhoul, "Introducing disparity map for monitoring and surveillance in wireless video sensor networks," in *International Conference on Computational Science (ICCS 2016)*, 2016.
- [16] W. Boonsuk, "Investigating effects of stereo baseline distance on accuracy of 3d projection for industrial robotic applications," in *5th IAJC/ISAM Joint International Conference*, November 2016.
- [17] P. Si, C. Wu, Y. Zhang, Z. Jia, P. Ji, and H. Chu, "Barrier coverage for 3d camera sensor networks," *Sensors*, vol. 17, no. 8, p. 1771, Aug 2017. [Online]. Available: <http://dx.doi.org/10.3390/s17081771>
- [18] H. Kim and A. Hilton, "3d scene reconstruction from multiple spherical stereo pairs," *Int. J. Comput. Vision*, vol. 104, no. 1, pp. 94–116, Aug. 2013. [Online]. Available: <http://dx.doi.org/10.1007/s11263-013-0616-1>
- [19] R. Dai and I. F. Akyildiz, "A spatial correlation model for visual information in wireless multimedia sensor networks," *IEEE Transactions on Multimedia*, vol. 11, no. 6, pp. 1148–1159, Oct 2009.
- [20] J. Delon and B. Rougé, "Small baseline stereovision," *Journal of Mathematical Imaging and Vision*, vol. 28, no. 3, pp. 209–223, Jul 2007. [Online]. Available: <https://doi.org/10.1007/s10851-007-0001-1>
- [21] D. Scharstein, H. Hirschmiller, Y. Kitajima, G. Krathwohl, N. Nescic, X. Wang, and P. Westling, "High-resolution stereo datasets with subpixel-accurate ground truth," in *GCPR*, ser. Lecture Notes in Computer Science, vol. 8753. Springer, 2014, pp. 31–42.