Efficient and accurate monitoring of the depth information in a Wireless Multimedia Sensor Network based surveillance

Anthony Tannoury and Rony Darazi
TICKET Lab.
Antonine University
Hadat-Baabda, Lebanon.
Email: anthony.tannoury@ua.edu.lb

Christophe Guyeux and Abdallah Makhoul
Femto-ST Institute, UMR 6174 CNRS
Université de Bourgogne Franche-Comté
France

Abstract—Wireless Multimedia Sensor Network (WMSN) is a promising technology capturing rich multimedia data like audio and video, which can be useful to monitor an environment under surveillance. However, many scenarios in real time monitoring requires 3D depth information. In this research work, we propose to use the disparity map that is computed from two or multiple images, in order to monitor the depth information in an object or event under surveillance using WMSN. Our system is based on distributed wireless sensors allowing us to notably reduce the computational time needed for 3D depth reconstruction, thus permitting the success of real time solutions. Each pair of sensors will capture images for a targeted place/object and will operate a Stereo Matching in order to create a Disparity Map. Disparity maps will give us the ability to decrease traffic on the bandwidth, because they are of low size. This will increase WMSN lifetime. Any event can be detected after computing the depth value for the target object in the scene, and also 3D scene reconstruction can be achieved with a disparity map and some reference(s) image(s) taken by the node(s).

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

I. INTRODUCTION

Monitoring and surveillance attract many researchers, specially with the development of the Internet of Things (IoTs). Within this discipline, Wireless Senor Networks (WSNs) are able to sense scalar data like temperature, light, and so on [1], while Wireless Multimedia Sensor Networks add the access to audio, images, or video data. They are built with low cost [2] Complementary Metal Oxide Semiconductor (CMOS) cameras and have a big range of applications in different fields like health, military, environmental, etc. The main challenges in WMSNs is to capture and process needed information with low energy consumption, fast computation time, and high Quality of Service (QoS).

The depth of an object or a monitored character gives us an essential clue for tracking or triggering an event. This depth will create a 3d representation of any object and it cannot be supplied by conventional gathered data. Most existing 3D reconstruction methods [3] use professional, complex, and expensive sensors with a specific network topology to obtain the required results. However, camera sensors in WMSNs have low energy resources because they are powered by limited batteries, so energy consumption is a primary constraint for WMSNs as we aim to prevent from decreasing uselessly the network lifetime. Our main objective is thus to compute 3D-depth while maintaining low power consumption, high speed, and QoS in the specific context of wireless multimedia sensor networks.

This is why we propose the use of *low-complexity* disparity maps computed by a couple of *low cost* sensor nodes in a WMSN. This is done to efficiently provide, to the monitoring system, an important depth information of the object under surveillance. Indeed, having multiple images captured at low cost from different views, and applying Stereo Vision on different couples after regrouping each pair of sensors as left camera and right one, our system can calculate at low cost some disparity maps to recover the depth information.

This system is summarized in Fig. 1. As can be seen, data processing is made locally in each couple of sensors. This will reduce the transmission rate that has a direct and important effect on the network lifetime. Our system will thus not transmit high resolution images, but only gray-scale disparity maps, and on demand – depending on the triggered event or changes in the scene.

The remainder of this article is organized as follow. Section II introduces the WMSN specifications, challenges, and limitations. In Section III, two low cost disparity map extraction methods are presented, namely the Sum of Absolute Differences (SAD) and Sum of Squared Differences (SSD) ones. They are experimented in the next section, and a complexity estimation is provided. This research article ends by a conclusion section, in which the contribution is summarized and intended future work is outlined.

II. WIRELESS MULTIMEDIA SENSOR NETWORKS

A. Composition of WMSNs

WMSN are primarily built with wireless sensors able to record images, audios, and videos. They are powered by batteries and they have the capability of sensing, processing,
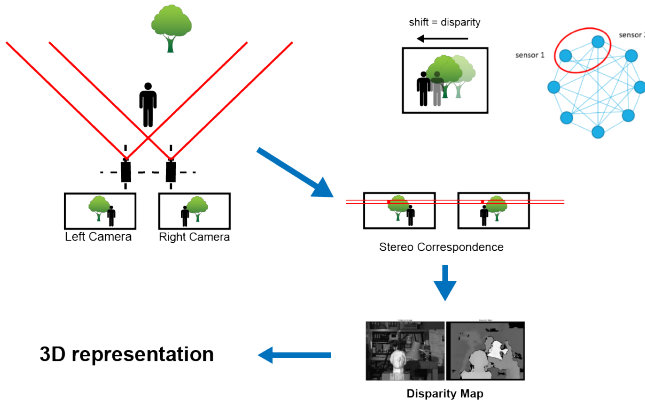
Fig. 1: Phases of our system

and transmitting data [4]. Finally, the availability of low cost CMOS cameras and the development of low complexity signal processing technologies and algorithms gives to the WMSNs the ability to provide useful multimedia data at low cost.

B. Applications of WMSNs

WMSNs can be used in various applications for divergent fields like object tracking, agricultural monitoring, e-health, and so on [5]. For instance, it can be deployed in a scene to monitor visible and hidden objects from different angles. Ham *et al.* [6] and Golparvar-Fard *et al.* [7] supervised buildings and constructions. Augmented reality can be mixed too with WMSNs to visualize real time some 3d representations using mobile software [8].

C. WMSN Challenges

The main constraint in WMSN is energy consumption, because sensors are powered by small batteries. So optimizing energy is very essential in our work independently from the type of application. Most of the energy is consumed by transmission. So decreasing the transmission rate and distance between nodes will increase network lifetime. As an illustration, 64KB data processed in a wireless sensor node leads to a consumption of $0.00195\mu J$ for program execution (data processing), which has no comparison with the $377\mu J$ of power required for data transmission (radio), as experimented in [9].

WMSNs need higher data rate while using high resolution images. Capturing images in a short time can be achieved but video streaming requires continuous capturing and delivery. In some monitoring scenarios, the system should work on real-time. Hence, powerful hardware and software techniques are a must to deliver the QoS requested by the considered applications.

D. Problematic and Solution

Our main problematic is how to develop an efficient WMSN-based monitoring that runs for a long period of time, works in real-time, while guaranteeing a good QoS [2]? In the studied context, the transferred data will be operated later for

3D scene reconstruction, at sink level. Stereo matching is done on each pair of sensors in order to deliver low size disparity maps saving depth information. Our system uses distributed WMSN, so the disparity map processing is shared recursively on all sensor nodes during the transfer to the sink, in order to decrease network load and energy consumption.

The next section outlines two low cost disparity map calculation methods that can reasonably be applied on WMSN.

III. DISPARITY MAP CALCULATION ON WMSN

Disparity map shows the pixels difference or motion between two stereo images captured from two (left and right) sensors. Being an hot topic, new developments and techniques are introduced each year to improve the quality of the produced maps. We are however not focusing on the most up-to-date methods of high quality, as they are complex, but we intend to choose the most appropriate disparity map method in the WMSN context. In other words, we do not target a disparity map of the best quality, but the best compromise between quality of this latter and complexity to obtain it.

Existing methods can be classified in two main categories, namely the local methods and the global ones.

In our previous work [10], we have investigated all reputed disparity map calculation methods. And we have chosen the sole local methods as unique convenient solutions for our WMSNs context, in terms of low complexity, fast speed, and reasonable quality. It is indeed well known [11] that global methods achieve good results, but are computationally expensive. Besides, local methods, especially the Sum of Absolute Differences (SAD) and Sum of Squared Differences (SSD) ones, bring off reasonable results while being low cost.

In SAD algorithm, the absolute difference between the intensity of each pixel in the reference block and the one of the corresponding pixel in the target block is computed with the following formula:

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

SAD makes the sum of differences over w, where w is the aggregated support window. The SSD algorithm, for its part, can be summarized as follows, see Equation 2:

$$SSD(x, y, d) = \sum_{(x,y)\in w} |I_l(x, y) - I_r(x - d, y)|^2 \quad (2)$$

We now intend to extend the aforementioned study by a deep experimentation and complexity calculation, to choose a good trade-off between speed, computation cost, and disparity map quality. The experiment and complexity computations, provided in the next section, ensure how Sum of Absolute Differences (SAD) is better than Sum of Squared Differences (SSD) with respects to computation time and complexity.
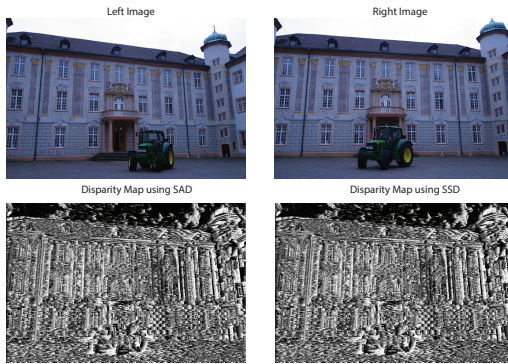
Fig. 2: Different views of the scene
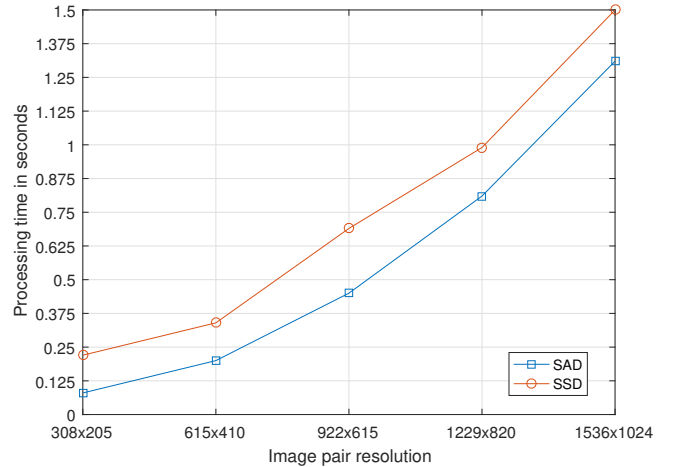

Fig. 3: Disparity Map using SAD and SSD

Pair	SSIM	PSNR
(node 1, node 2)	1.0	$\approx infinity$
(node 1, node 3)	1.0	$\approx infinity$
(node 1, node 4)	1.0	$\approx infinity$

TABLE I: SSIM and PSNR

(a) Processing Time with different image resolutions

(b) Processing Time with different coupling
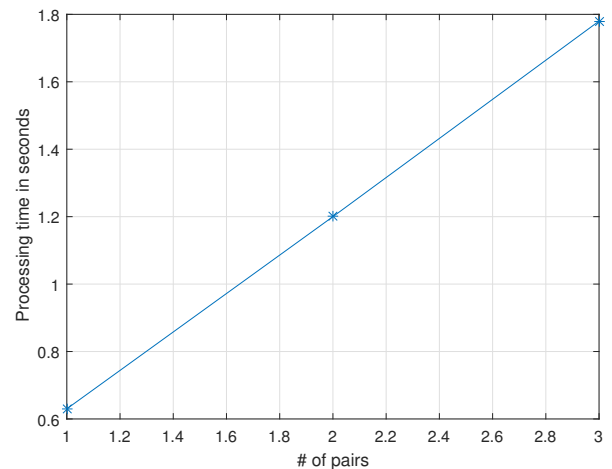
(c) Processing Time with different number of pairs

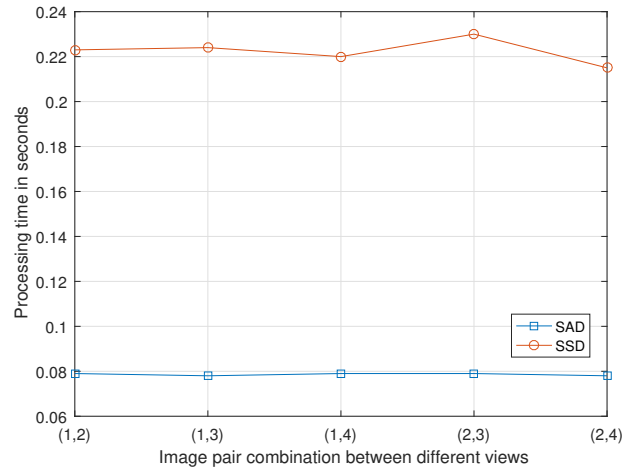Fig. 4: Processing time per resolution, views and number of pairs.

IV. EXPERIMENTS AND COMPLEXITY COMPUTATION

The Strecha *et al.* [12] multi-view dataset has been used for comparison, while all simulations are computed using Matlab 2016. Fig. 2 shows different views of the same scene. Disparity maps using SAD and SSD are depicted in Fig. 3. Fig. 4a emphasizes that SAD is better than SSD when focusing on time performance, while the computation time decreases with the captured images resolutions. Note that the performance is independent from sensor coupling. This is obvious in Fig. 4b, where the processing time is approximately the same for different views. Finally, Fig. 4c shows how processing time increases linearly with the number of chosen couples.

Table I contains structural similarity (SSIM) and Peak Signal-to-Noise Ratio (PSNR) functions applied on different pairs. SSIM and PSNR are used to measure the similarity between the two calculated disparity maps (using SAD or SSD). Getting 1.0 and infinity for different pairs ensure that the calculated disparity maps have approximately the same quality and are too similar. This will help us later to choose

SAD as the best method – because it has the same quality as SSD, but with a lower complexity leading to a more efficient processing.

Indeed, the complexity of the two methods is accessible theoretically, proving that SSD is more complex than SAD. In this first situation, WMSNs will need more time to process more complex algorithms, and then it will consume more energy [13]. Computing the complexity of the two approaches using Equations 1 and 2, we have obtained the following results. For the SAD, we have 2 loops, one for horizontal width and the second one for vertical height of the image, while the operation within the loop is a single subtraction. So the SAD complexity is $O(n^2)$ in terms of elementary operations. The SSD, for its part, integrates a square within the same loop, increasing the complexity to $O(n^3)$ elementary operations, where n is the number of lines (or columns) in the images. Such complexities are coherent with the simulations, leading to the choice of SAD as best compromise to achieve an acceptable depth evaluation in a WMSN-based video surveillance context.

V. CONCLUSION AND FUTURE WORKS

Disparity map is a main parameter to get the depth of a monitored scene and then reconstruct it. In this research article, we contributed by applying disparity map calculation on WMSN distributed nodes. Our approach is directed by the main WMSN limitations: energy consumption, processing capability, QoS, and communication. The experiments and studies we done help to choose the best disparity map calculation method for WMSN usage, namely the SAD approach.

In future work, we intend to investigate not the existing literature, but novel disparity map computation methods specifically designed for wireless multimedia sensor networks. They will reach the optimized compromise between quality of the maps and complexity to obtain them. We will further investigate the optimal way to transfer the map from the terminal couple of nodes until the sink, by updating the disparities at aggregator nodes when receiving maps from various close couples of sensors (that observe a similar scene). The proposal will finally be distributed in a real wireless sensor networks, in order to test *in vivo* the real performance of this optimized surveillance in operational context.

REFERENCES

[1] A. Sharif, V. Potdar, and E. Chang, "Wireless multimedia sensor network technology: A survey," in *2009 7th IEEE International Conference on Industrial Informatics*, June 2009, pp. 606–613.

[2] N. Boudargham, J. B. Abdo, J. Demerjian, C. Guyeux, and A. Makhoul, "Investigating low level protocols for wireless body sensor networks," in *13th ACS/IEEE International Conference on Computer Systems and Applications (AICCSA)*, vol. *, 2016.

[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, 2014. [Online]. Available: http://dx.doi.org/10.1007/s11042-012-1184-z

[4] I. T. Almalkawi, M. Guerrero Zapata, J. N. Al-Karaki, and J. Morillo-Pozo, "Wireless multimedia sensor networks: Current trends and future directions," *Sensors*, vol. 10, no. 7, p. 6662, 2010. [Online]. Available: http://www.mdpi.com/1424-8220/10/7/6662

[5] 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*. Springer Publishing Company, Incorporated, 2013.

[6] Y. Ham, K. K. Han, J. J. Lin, and M. Golparvar-Fard, "Visual monitoring of civil infrastructure systems via camera-equipped unmanned aerial vehicles (uavs): a review of related works," *Visualization in Engineering*, vol. 4, no. 1, pp. 1–8, 2016. [Online]. Available: http://dx.doi.org/10.1186/s40327-015-0029-z

[7] M. Golparvar-Fard, F. Pea-Mora, and S. Savarese, "Monitoring changes of 3d building elements from unordered photo collections," in *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on*, Nov 2011, pp. 249–256.

[8] D. Goldsmith, F. Liarokapis, G. Malone, and J. Kemp, "Augmented reality environmental monitoring using wireless sensor networks," in *2008 12th International Conference Information Visualisation*, July 2008, pp. 539–544.

[9] 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.

[10] A. Tannoury, R. Darazi, A. Makhoul, and C. Guyeux, "Introducing disparity map for 3d scene reconstruction in wireless multimedia sensor networks," pp. ***–***, December 2016, submitted article.

[11] K.-J. Yoon and I. S. Kweon, "Adaptive support-weight approach for correspondence search," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 4, pp. 650–656, 2006.

[12] C. Strecha, W. von Hansen, L. V. Gool, P. Fua, and U. Thoennessen, "On benchmarking camera calibration and multi-view stereo for high resolution imagery," in *2008 IEEE Conference on Computer Vision and Pattern Recognition*, June 2008, pp. 1–8.

[13] G. Mao, B. Fidan, G. Mao, and B. Fidan, *Localization Algorithms and Strategies for Wireless Sensor Networks*. Hershey, PA: Information Science Reference - Imprint of: IGI Publishing, 2009.