The 2010 IEEE/RSJ International Conference on
Intelligent Robots and Systems
October 18-22, 2010, Taipei, Taiwan

Humanitarian Demining Robot Gryphon: New Vision Techniques and Optimization Methods

Alex Masuo Kaneko, Marco Marino and Edwardo F. Fukushima

Abstract— **Landmines continue killing and injuring people even many years after the end of conflicts. Mechanical systems and robots have been used for assisting in landmines location and improving the quality of humanitarian demining tasks. These new systems are expected to enhance safety, efficiency, time duration, etc, but many improvements are still needed in order to realize a practical system. In this paper, new vision techniques for improving the effectiveness of terrain mapping using stereo vision camera under high-contrast light conditions, i.e., mixed direct sunlight and shadowed areas, are introduced. Furthermore, trajectory optimization taking in account the obtained terrain map is also analyzed, which greatly reduces time consumption for landmine location marking tasks. These newly introduced methods greatly contribute to overall increase in reliability and efficiency of humanitarian demining tasks.**

I. INTRODUCTION

Landmines are a plague that has existed for many decades. While the international community is largely cooperating in preventive action (in trying to ban the use and spread of those weapons), efforts are also being focused in remedial action (in neutralizing buried landmines, in assisting affected individuals and populations). Landmine neutralization is dangerous, costly and tedious. Tools are still rudimentary and the most commonly used landmine sensors still suffer from high rates of false positives [1].

The Tokyo Institute of Technology developed a semi-autonomous mobile robot called Gryphon-V (shown in Fig.1), to assist the mine detection process. Its manipulator is able to automatically scan over rough terrain, record data and present the resulting sensor images to the operator who then can mark suspect spots. The developed robot has been tested in several field trials on test minefields in Croatia and Cambodia, showing promising results [2].

Several other attempts have been made in the world, in automating or assisting human deminers in the scanning process; legged robots [3], wheeled vehicles, tracked vehicles and even suspended inspection tools [1] have been researched. Unfortunately, research is often focusing on one particular aspect (e.g. locomotion or sensing) leading to weak system integration which results in an ineffective and slow demining operation.

For the Gryphon system in particular, most attention has been paid in the mechanical and electrical design to build a rugged system, and also main consideration of sensing imaging for maximization of POD (Probability Of Detection) and minimization of FAR (False Alarm Rate) for the mine sensing tasks. This proved to be effective, and the latest tests in Croatia [2] showed that the Gryphon system has the potential to be "as good as humans or even better" considering POD and FAR indexes. However, other important factors to consider are: i) "time for each operation", and the ii) influence of "human factors".

In this article, the "time for each operation" factor is considered, and methods for improving the overall efficiency of the system is proposed, including new vision techniques and trajectory optimization methods.

II. GRYPHON DEMINING TASKS

The basic tasks for the demining process using Gryphon are:

1. Start of day procedure: performs auto-checking of system components and move the manipulator from the resting position to the standby position, and brings the system to be ready to start the demining operation.
2. Mapping: the depth map, i.e, the 3D profile, of the area to be demined is mapped by use of a stereo vision camera.
3. Reference points/line setting: the area to be scanned is selected by a group of points/line.

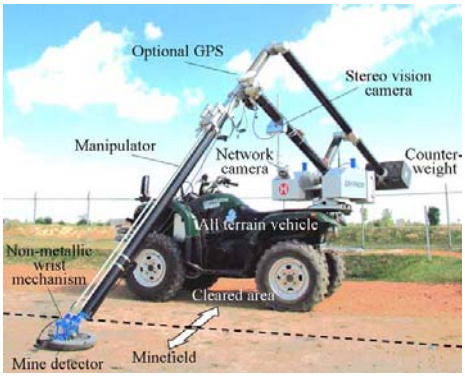

Fig 1 – Overview of Humanitarian Demining Robot Gryphon

A. M. Kaneko, M. Marino and E. F. Fukushima are with the Department of Mechanical and Aerospace Engineering, Tokyo Institute of Technology, Japan. alex@robotics.mes.titech.ac.jp

4. Scanning: the manipulator is moved inside the scan area and the Metal Detector (MD) signal is captured.
5. Marking: suspect mines locations are painted/marked.
6. Moving to the next area: the robot is moved to the next area to be scanned.
7. End of day procedure: the manipulator is brought to the resting position, and the robot is ready to return from the minefield.

The "mapping" and "marking" tasks are described in more detail in the following sections.

2.1 Mapping Task

The mapping task is responsible for capturing the environment information, where the robot performs the landmines detection. For this purpose, a stereo vision camera fixed to the robot arm is used (Fig. 1). Gryphon-V in particular, is equipped with Point Grey Bumblebee™. During the mapping task the camera is oriented to different angles by moving the arm to different positions (P1, P2, P3), thus enabling capturing information of separate areas (A1, A2, A3) that can be concatenated to one larger area (Fig.2) to be scanned.

2.2 Mapping Task Problems

Although the commercial stereo vision camera used in Gryphon showed good performance in most outdoor environments during field tests, it has been observed that for some extreme lighting/shadow conditions, the stereo correspondence algorithm is unable to find enough features to perform a depth analysis, and the mapping task cannot be successfully finished.

The problem is mainly observed when there are regions of strong sunlight and shadow in the same picture. For this case, some areas of the image saturates while shadowed areas appears as dark images. Such extreme case of lighting is not uncommon in the field, and even the shadow of the robot itself can cause problems to the depth map calculation.

Next section introduces new vision techniques to deal with this problem.

III. New Vision Techniques

Finding correspondences between features is the key of stereo processing, but in some undesirable cases, the number of features can be extremely low to perform any kind of analysis.

For instance, consider Fig. 3. It shows a pair (right and left lens of the camera) taken by Gryphon's stereo vision camera in laboratory environment, using a portable version of the manipulator. The extreme lighting environment simulated direct sunlight by using a high power directional light. The figure also shows the depth map calculated from the stereo pair and displayed in the user interface. As can clearly be seen it lacks so many features that only a fraction of the image was actually used to compute 3D information.

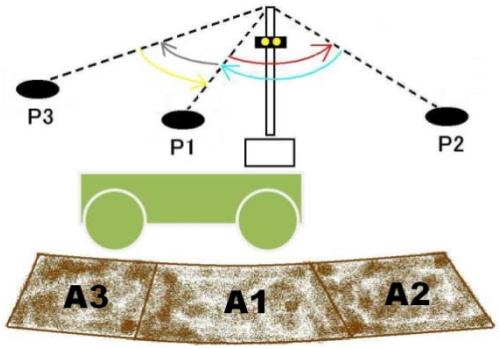

Fig. 2 –Main steps of the mapping task

The stereo correspondence algorithm is not imputable for such an undesirable behavior, because for limits of the technology itself, it would be impossible to efficiently compute the depth map when features are lacking. The true limit in this case is coming from the camera and its limited dynamic range.

This section describes the vision techniques implemented in this work, based on the High Dynamic Range (HDR) and Exposure Fusion (EF) techniques.

3.1 Existing Techniques

A camera is not able to represent a large diversity of light conditions in one single picture, but it can obtain a good result by combining information coming from several pictures. As the camera can alter, through driver control, the behavior relative to the exposure control, it is possible to obtain visual information on the areas that would otherwise be thresholded either to zero (black) or to the maximum (white) level. The idea behind the approach is then to take multiple pictures and combine them together. Two techniques will be analyzed, one being HDR and the other one EF.

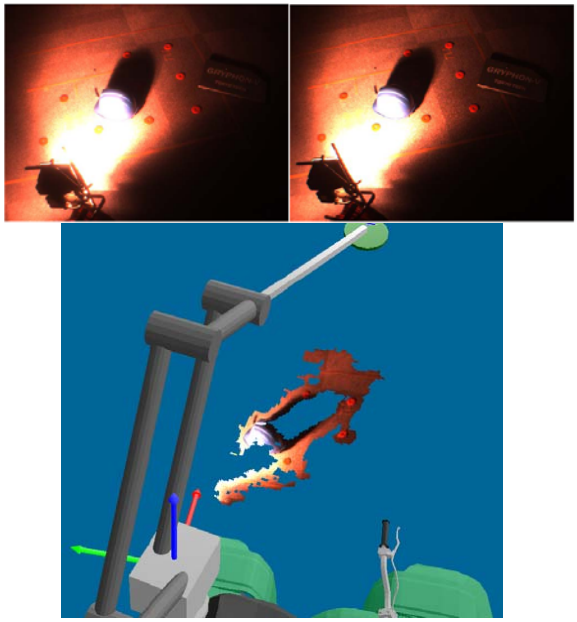

Fig. 3 – Effect of strong light: real and virtual cases

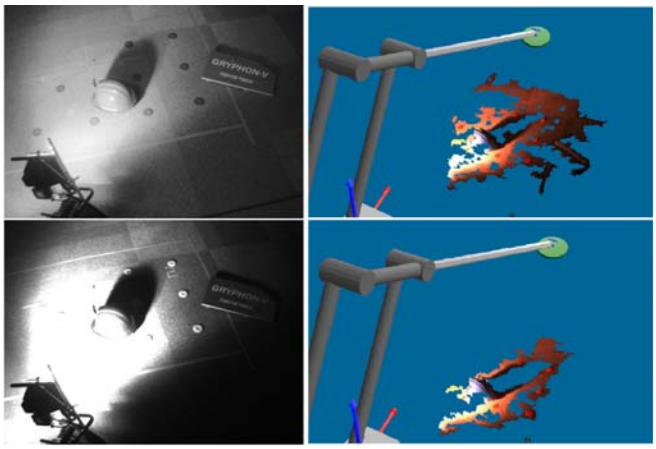

Fig. 4 – Real and virtual cases: before (down) and after (top) the technique implementation

HDR technique is well-known in the digital photography environment, and gives the overall best result, but it is mathematically complex, what makes it difficult to be implemented in real-time. Conceptual notions from HDR are basically coming from the first US patent described in [4] with a major influence from the work of Debevec in 1997 [5].

EF basics are introduced in the same fashion the author describes in his original paper [6]. The idea behind the approach is to compute a perceptual quality measure for each pixel in the multi-exposure sequence, encoding desirable qualities like saturation and contrast; in this way it is possible to select the "good" pixel in each image which is the candidate to form the final output image. As HDR, this technique also requires the perfect alignment of the input images.

3.2 Proposed Implemented Vision Technique

The algorithm proposed and implemented in this research, aims at reaching a good level of fusion, taking HDR as the maximum possible qualitative achievement, but following an approach more similar to EF.

The algorithm takes, as EF does, the average image by considering some weights when balancing the pictures. The main difference with respect to the known algorithm [6], is that the weights work in a discrete logical fashion.

The function used to compute the weights is based on the average brightness (luminance) level of each image. This level should move over a certain threshold, which is defined at the startup of the algorithm by a statistical training procedure, taking several pictures with the stereo camera in many different possible light conditions. A common value was obtained, which statistically proved, in the considered example, that at least 80% of the image was saturated to the maximum reading level, corresponding to white. The algorithm used the threshold level to set a 1.0 weight only for images which presented at least 20% of valid information, with respect to the total data.

This particular behavior of the weight computation function allows to automatically exclude overexposed images, as well as allows the algorithm to bias the search for a solution with respect to the lower exposure pairs. Instead of using three different sources of information as classical EF is doing, the luminance information is the only one eventually used, resulting in a speed-up of the overall computation time.

Another difference is that the algorithm always works for every stereo image pair as input, instead of dealing with only one image at a time.

Several configurations were experimented to choose the right number of pictures to be taken, and the resulted quantity was 6 different exposures. This particular choice was accurately made by considering the exposures range of the camera, which ranged from EV 0 to EV 1000. This stepping of EV 200 was considered enough to avoid information loss between one step and the other one. A total of 6 steps are naturally calculated as (EV 1000 / EV 200) + 1, with the unity added to take into account the EV 0 factor.

In mathematical terms, the pseudo-mean image is computed in a fashion similar to EF, that is:

$$O_{i,j} = \sum_{k=1}^{v} W_k . I_{i,j,k} \qquad (1)$$

The equation states that the pixels ij in the output image O are computed as an average mean weighted for each input image k by the dynamic weight W. The v term on top of the summation symbol is a variable entity, depending on the result of the weighting function for each image k.

While the result would be the same as having v = N, (N = maximum number of images), the ability to limit it to a possibly smaller value (v < N), can speed up the combining process, since we do not need to multiply the terms inside the summation by a zero weight hundreds of times.

As can be seen in Fig. 4, the performance improvement is traded with the contrast of the image, but in some areas of the image, like in the bottom-right part, the quality improvement is more evident.

The overall contrast is in average much lower than the two analyzed standard techniques, leading to a proximity to the average brightness level, which is $\frac{B_{min} + B_{max}}{2}$, where "B" simply stands for brightness. However, this lack in the vivacity in colors (or grey levels) is not really important for stereo vision, because it is enough to improve the number of features within the image. An example of application for the algorithm in normal light conditions can be verified in Figs 5 and 6.

As can clearly be seen in the sample pictures, the algorithm performance surpasses the standard procedure taken by the camera even with regular light conditions.A more analytical comparison between the two image pairs was done (Table I) using SURF.feature extraction algorithm [7], which permits counting the overall number of features detected. Here it is only used as a quantitative measuring device, useful to estimate the number of "good" features present in an image pair.

It is possible to infer that the increase in pair shown in Fig. 4, which is also the one strongly affected by extreme lighting condition, is very high (almost 60%). In Fig.6, under standard lighting condition, the increase is even slightly negative, less than 2% of difference can be confidently considered a minimal difference.

Experimental results suggest that the function can be used in every light condition, because it can offer similar results during normal lightning, and boosts as far as 60% the overall quality of the image in extreme lighting conditions.

Fig. 7 shows graphically the SURF features computed on the pair in Fig. 6, in order to show how many and where the features are detected.

3.3 Alignment with SURF

Vibrations coming from the wind or the engine of the All Terrain Vehicle (Fig.1) reduce the precision of the robot. Little vibrations can propagate through the long arm, interfering during the demining operation. In case of the camera, vibrations make the taken pictures become blurred, leading to potential problems, especially for the proposed methods (that require six steady pictures).

For solving the alignment problem an algorithm was proposed, which is based on a random correlation window, which computes SURF features, in both images, that are later compared to each other, in order to find the one and single feature which is most similar between them. In this way we can be confident that the chosen point to calculate the misalignment corresponds to the same physical points in both images. An experiment done in laboratory, which the robot's arm was manually heavily shaken is shown in Fig. 8.

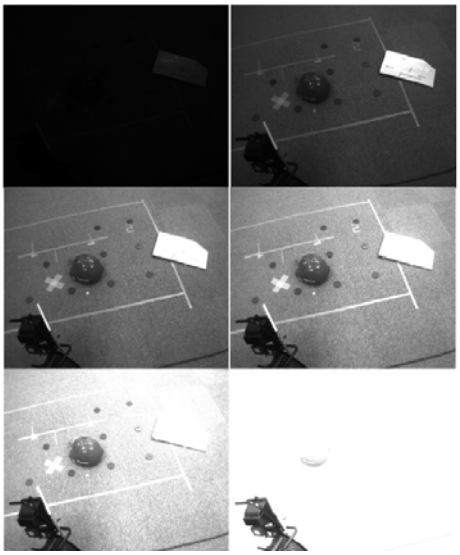

Fig. 5 – Source images taken in regular light conditions

TABLE I.
Comparative results

Pair	Standard	Compensated	Increase
Fig. 4	2493	3920	+54,24%
Fig. 6	5657	5551	-1,87%

IV. TRAJECTORY OPTIMIZATION IMPLEMENTATIONS

This section describes further optimization methods for the marking trajectory and for the marking sequence with Gryphon that can be reliably implemented after the enhancements in vision.

4.1 Trajectory Optimization

The marking task with Gryphon can be described in four main steps (Fig. 9): button pressing in the interface (a), motion from P1 to P2 (b), mark painting (c) and motion from P2 to P1 (d). The described steps are performed for each mark. However, according to the terrain, many marks may be needed, and the described procedures will be repeated as many times as the input number of marks.

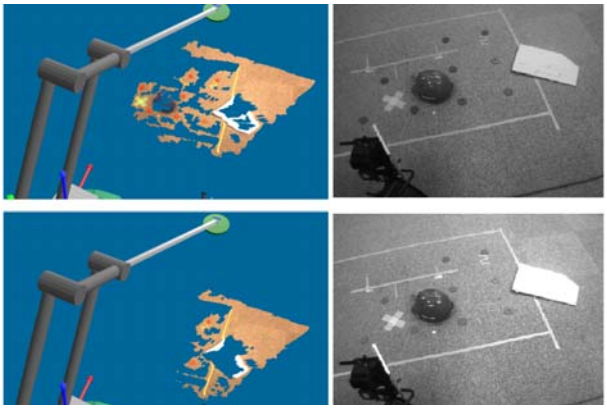

Fig. 6 – Compensated (top) and original (bottom) image

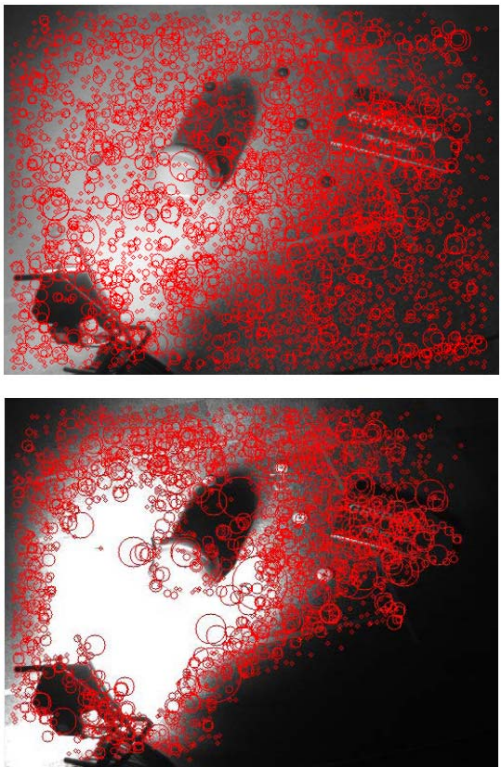

Fig. 7 – Visual comparison of SURF features of pair in Fig. 5

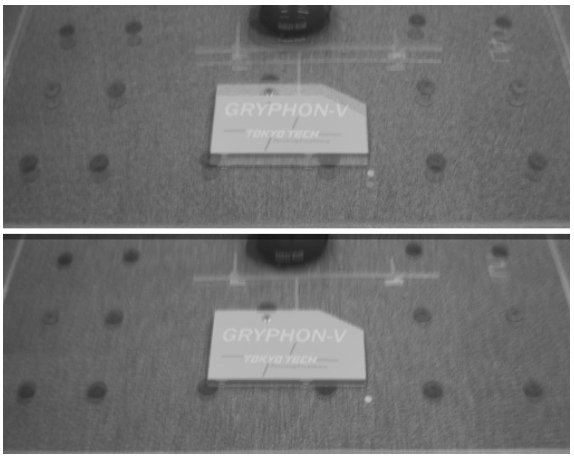

Fig. 8: Example of alignment result (bottom) and original image (top)

For example, if the robot has to draw two marks, the robot's arm will do the following trajectory: P1, P2, P1, P3 and P1.

Regarding the shown trajectory problem, the main requisite is that the arm does not go back to the initial position (P1) and goes directly to the next spot. The problem is that the ground usually has many variations in height, and if the terrain information is not correctly grabbed due to vision problems, the arm could easily hit an obstacle.

Once the terrain containing the two points is not flat, a simple straight line cannot be used as a trajectory and an adequate (as short as possible) path has to be generated. Many previous works were performed to solve this problem [8-11], but here as a first approach a simpler algorithm was applied.

Consider Fig. 10. The proposed method consists of building a line (shown in yellow) between the actual point (A) and the next point (B). Since we know the coordinates of A (X1, Y1) and B (X2, Y2), we can simply substitute them in the line equation:

$$y = a \,.\, x + b \tag{2}$$

where "*a*" is the angular coefficient, defined as:

$$a = \frac{Y_2 - Y_1}{X_2 - X_1} \tag{3}$$

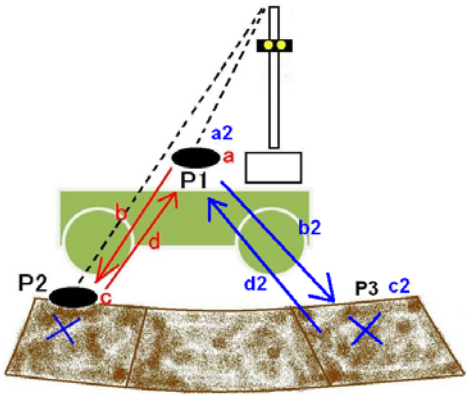

Fig. 9 – Trajectory between two consecutive marks

and "b" is the linear coefficient, defined as:

$$b = Y1 - a \,.\, X1 = Y2 - a \,.\, X2 \tag{4}$$

Once we have this line, we have a plane perpendicular to the ground (in green), that will guide the arm until the next point. Finally, the robot calculates the height fields (Z coordinates, precisely obtained with the new techniques with the stereo vision camera) in this plane, and moves the arm through these height variations in the terrain.

4.2 Sequence Optimization

After proposing an optimized solution for the marking trajectory, another optimization became evident. Since the robot paints the marks in the inverse order they were input in the interface by the operator, in case we have three or more marks, an optimized sequence should be considered.

Since asking the operator to input the marks in an optimized sequence would require too much workload and long decision making times, a more reasonable approach was adopted, which is the robot itself computes and chooses the shortest path to go through all the points. Here, many previous works such as [12] were done to find the shortest path passing through all the points. However, this algorithm is complex to be implemented and as a first optimization proposal, an algorithm similar to [13] was implemented.

Suppose the operator has set the following sequence in the interface: "1","2","3","4","5","6","7" and "8". With the proposed method, the robot gets the first mark of the sequence ("1") and calculates the distance between this first mark and the remaining ones. For this first approach, only coordinates X and Y are being used, since variations in Z are usually relatively smaller.

In the example shown in Fig. 11, the robot finds that the nearest mark from "1" is "3", and uses this last mark as verified. The robot calculates the total distance obtained with this path having mark "1" as initial point and stores this value.

Then, the robot sets the next mark ("2") of the original sequence as initial point and repeats all the procedures. These procedures are repeated until the final mark ("8") is set as initial point.

After calculating the total distance of all possible paths (eight in this example), the robot verifies which of them has the shortest path, and reorganizes the original sequence into this optimized one (Fig. 12).

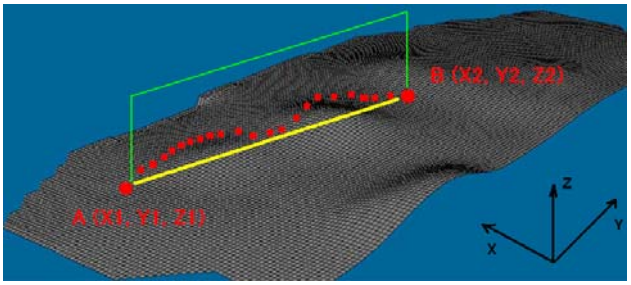

Fig. 10 – Trajectory optimization

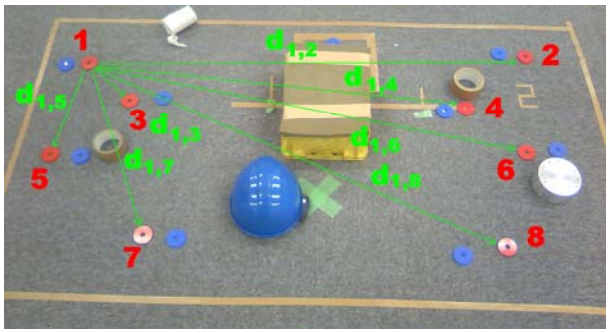

Fig. 11 – Sequence optimization

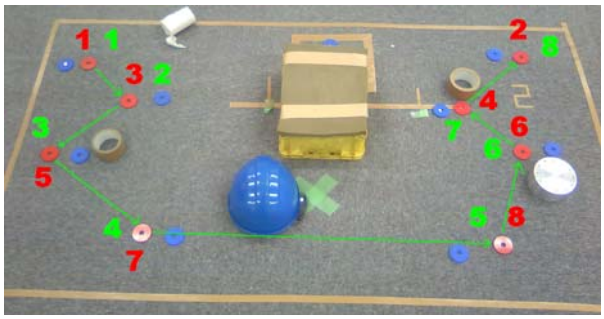

Fig. 12 – Sequence optimization test

TABLE II.
Tests results

Original (s)	Trajectory Optimization (s)	Trajectory and Sequence Optimization (s)
235 s	141 s	120 s

4.3 Optimizations Evaluation

Considering the same configuration (number and position of the marks) in Figs 11 and 12, a test was performed to evaluate the magnitude of the time duration between the methods. From Table II, we can observe that the time duration was hugely improved after applying both optimizations, compared to the original marking system (about 49%). Here, it is important to notice that the time duration is function of the number of marks, distance, input sequence and operator performance. In other words, variations in these variables can lead to different results, but it can be said that in the worst case, the time duration of the optimized system will be similar to the previous one.

V. CONCLUSIONS

A new technique of a HDR-like for real time stereo vision was developed in this work. With the presented technique, it is possible to improve the number of features up to 60% with respect to the original image pair. Another aspect is about the use of SURF features for fast realignment of image pairs.

With these vision techniques, the terrain map can be effectively done under unfavorable light conditions, improving the reliability of the demining operation.

Optimization methods for improving the times and decreasing the operator workload in demining operations with Gryphon were also presented, in particular for the marking sequence determination. Although the applied algorithms do not guarantee the most efficient results (according to works existing in literature), both trajectory and sequence optimization proved to be necessary and showed satisfactory results.

Furthermore, the introduced vision enhancements and trajectory optimization techniques can also be applied to a variety of systems other than Gryphon.

VII. FUTURE WORKS

Further optimizations in Gryphon system are being considered. Time duration is an important factor in demining operations, but another crucial one is the human factors. The actual system still requires considerable work load from the operator during the landmine discrimination and in the overall operation. The user interface is being improved to increase the level of automation of each operation/task, resulting in a faster and more reliable system to be used in practical situations.

REFERENCES

[1] E. F. Fukushima, M. A. Freese, T. Matsuzawa, T. Aibara and S. Hirose. "Humanitarian Demining Robot Gryphon: Current Status and Objective Evaluation". International Journal on Smart Sensing and Intelligent Systems, Vol. 1, No. 3. September 2008.

[2] Nikola Pavkovic, Jun Ishikawa, Katsuhisa Furuta, Kazunori Takahashi, Mate Gaal, Dieter Guelle, "Test and Evaluation of Japanese GPR-EMI Dual Sensor Systems at Benkovac Test Site in Croatia", HCR-CTRO_TECH_GPR 08-001, March, 2008.

[3] S. Hirose, K. Kato, "Development of Quadruped Walking Robot With the Mission of Mine Detection and Removal", Proc. IEEE int. Conf. on Robot. and Automat., Leuven, Belgium, 1998, pp. 1713-1718.

[4] Wide dynamic range camera USA Patent 5,144,442. 1992.

[5] Paul E. Debevec, Jitendra Malik, Recovering high dynamic range radiance maps from photographs, International Conference on Computer Graphics and Interactive Techniques archive, Proceedings of the 24th annual conference on Computer graphics and interactive techniques table of contents, Pages 369 - 378, 1997

[6] Tom Mertens, Jan Kautz and Frank Van Reeth, Exposure Fusion, Pacific Graphics 2007.

[7] Herbert Bay, Andreas Ess, Tinne Tuytelaars, Luc Van Gool, "SURF: Speeded Up Robust Features", Computer Vision and Image Understanding (CVIU), Vol. 110, No. 3, pp. 346--359, 2008

[8] L. E. Dubins, "On curves of minimal length with a constraint on average curvature and with prescribed initial and terminal positions and tangents," Amer. J. Math., vol. 79, pp. 497?516, 1957.

[9] J. A. Reeds and L. A. Shepp, "Optimal paths for a car that goes both forward and backward," J. Pacific Math., vol. 145, no. 2, pp. 367?393, 1990.

[10] P. Soueres and J. P. Laumond, "Shortest paths synthesis for a car-like robot," IEEE Trans. Automat. Contr., vol. 41, pp. 672?688, May 1996.

[11] A. Bicchi, G. Casalino, and C. Santilli, "Planning shortest bounded-curvature paths for a class of nonholonomic vehicles among obstacles," J. Intell. Robot. Syst., vol. 16, pp. 387?405, 1996.

[12] Floyd, Robert W., "Algorithm 97: Shortest Path", Communications of the ACM, 1962.

[13] E. W. Dijkstra, "A note on two problems in connection with graphs", Numerishe Mathematik, 1959, pp. 269-271.