Application of Super-Resolution Algorithms for the Navigation of Autonomous Mobile Robots

Krzysztof Okarma, Mateusz Tecław, and Piotr Lech

West Pomeranian University of Technology, Szczecin
Faculty of Electrical Engineering
Department of Signal Processing and Multimedia Engineering
26. Kwietnia 10, 71-126 Szczecin, Poland
{krzysztof.okarma,mateusz.teclaw,piotr.lech}@zut.edu.pl

Abstract. In the paper the idea of using the super-resolution algorithms for the self-localization and vision based navigation of autonomous mobile robots is discussed. Since such task is often limited both by the limited resolution of the mounted video camera as well as the available computational resources, a typical approach for video based navigation of mobile robots, similarly as many small flying robots (drones), is using low resolution cameras equipped with average class lenses. The images captured by such video system should be further processed in order to extract the data useful for real-time control of robot's motion. In some simplified systems such navigation, especially in the within an enclosed environment (interior), is based on the edge and corner detection and binary image analysis, which could be troublesome for low resolution images.

Considering the possibilities of obtaining higher resolution images from low resolution image sequences, the accuracy of such edge and corner detections may be improved by the application of super-resolution algorithms. In order to verify the usefulness of such approach some experiments have been conducted based on the processing of the captured sequences of the HD images further downsampled and reconstructed using the super-resolution algorithms. Obtained results have been reported in the last section of the paper.

1 Introduction

Vision based autonomous mobile robot navigation becomes more and more interesting field of research during recent years. Growing interest in using real-time image and video analysis methods for the control of autonomous guided vehicles (AGVs) or unmanned aerial vehicles (UAVs) as well as many other robotic applications is caused mainly by the increase of the available computational power of processors and availability of relatively cheap cameras. Hence computer vision methods, together with cameras treated as complex passive sensors, can be considered as very attractive supplement to active or passive solutions typically used in robotics, such as infrared, laser or sonar sensors [1].

R.S. Choraś (ed.), *Image Processing & Communications Challenges 6*,
Advances in Intelligent Systems and Computing 313, DOI: 10.1007/978-3-319-10662-5_18

The advantages of vision systems in comparison to active sensors which alter the environment are well known, similarly as some limitations of GPS solutions, laser range finders or sonars (e.g. poor angular resolution). Therefore some applications of vision for the navigation of mobile robots have been considered even in 1980s and 1990s. Due to the development of cameras and video processing systems in consecutive years, some classical algorithms, often based on binary image analysis, have become more complex. Nevertheless, many solutions and algorithms are dedicated to a limited area of applications e.g. indoor navigation in corridors [2] or underwater navigation [3,4]. Some more advanced concepts may be based e.g. on saliency and visual attention [5] or Scale Invariant Feature Transform (SIFT) [6]. A comprehensive survey of many approaches can also be found in the paper [7].

Considering the types of navigation task (indoor or outdoor environment, map based, mapless or map building approach, structured or unstructured environments), the video analysis in low cost mobile robots is often conducted using simplified representation of images. Hence, the accuracy of edge and corner detection as well as their assignment in consecutive video frames is still a crucial element both in indoor and outdoor navigation.

For low cost autonomous robots (e.g. aerial ones), equipped with light low resolution cameras, high accuracy of self-positioning and navigation may be hard to achieve due to the physical limitations of the computer vision system. Nevertheless, especially in systems where a slight delay (less than a second) is acceptable, the accuracy of detection of some landmarks, edges or corners may be improved by using the super-resolution (SR) methods allowing the reconstruction of a single high resolution image based on a number of low resolution video frames assuming the presence of subpixel displacements in consecutive frames. Such situation is quite typical for a moving robot so the application of the super-resolution algorithms seems to be an interesting idea which can be useful for some of the mobile robot navigation systems.

2 Super-Resolution Algorithms

A typical approach to enlarging the digital images is based on the interpolation, using different methods e.g. nearest-neighbor, Lanczos resampling, bilinear or bicubic interpolation, supersampling or more advanced algorithms, such as patented S-Spline. Those algorithms are mostly based only on a single image in contrast to SR algorithms utilizing several consecutive video frames utilizing the sub-pixel shifts between multiple low resolution images representing the same scene, leading to the increase of the physical resolution of the output image.

During last several years numerous super-resolution algorithms have been proposed, including frequency domain algorithms and typically more computationally complex spatial domain methods which are less sensitive to model errors. Since a typical SR algorithm consists of two main steps related to image registration with motion estimation and image reconstruction, usually the latter is different in many algorithms. Image registration may be based on the analysis of the amplitude and phase frequency characteristics [8,9] or expansion into

Taylor series [10]. Some of the most popular approaches to image reconstruction algorithms are: iterated back-projection proposed by Irani and Peleg [11,12], Projection Onto Convex Sets (POCS), Papoulis-Gerchberg method [13], robust super-resolution [14] and normalized convolution [15].

Nevertheless, due to the computational demands of most of those methods, the bicubic interpolation combined with the image registration algorithm proposed in the paper [8] has been chosen. Considering also the possibilities of effective implementation of the algorithm in autonomous mobile robots, verification of the usefulness of the SR algorithms for the preprocessing of images used for mobile robot navigation purposes has been limited to algorithms with relatively low computational and memory demands.

3 Experimental Verification

The experiments have been conducted using two Full HD video sequences captured by the camera mounted on the Mobot-Explorer-A1 mobile robot. The first one has been captured in the corridor and the second one is an outside one. Such obtained frames have been considered as the reference and the video sequences have been downsampled by 3 in order to obtain low resolution frames being the input for the SR algorithm. For the comparison with the application of typical bicubic interpolation, the Structural Similarity (SSIM) metric [16] values have been calculated for the interpolated and the SR reconstructed images using the original Full HD frames as the reference.

Considering the fact that the most relevant areas of images for robot navigation purposes are located near the edges present in the consecutive video frames, a modification of the SSIM metric known as Three-Component SSIM (3-SSIM) [17] has been adapted for the additional verification of the impact of the SR algorithm on the obtained results. Calculation of the 3-SSIM metric is conducted separately for the three types of regions representing edges, textures and flat image areas. Since the most important regions in view of robot navigation are located near the edges the images have been masked by the result of morphological dilation of the edge detection result obtained using well known Canny filter. Then, the local values of the SSIM metric have been calculated only for those regions and finally averaged in order to obtain the final similarity score. The exemplary frames from both video sequences together with results of dilated edge filtering (shown as negatives) are presented in Fig. 1. As can be observed, the indoor frame is much less complex and contains smaller amount of edge information so the navigation task should be much simpler.

Since one of the relevant parameters of each super-resolution method is the number of input frames, all the experiments have been conducted for 3, 5, 7 and 10 frames used for the Full HD image reconstruction. Another important issue is related to the relative position of the reconstructed frame as better results are obtained for the middle frames from the sequence. Nevertheless, in order to reduce the amount of presented data, only the average values of the similarity scores calculated for each reconstructed frame are presented in Fig. 2. Almost

Fig. 1. Exemplary video frames for the indoor and outdoor sequences with results of Canny edge detection after dilation using 5 × 5 pixels structuring element

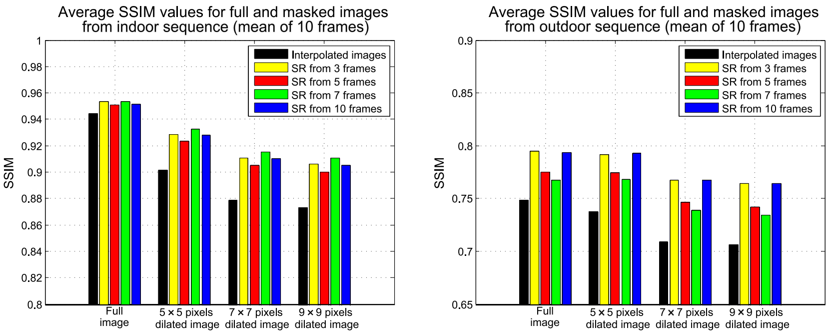

Fig. 2. Illustration of average SSIM and adapted 3-SSIM values calculated for full images and masked by dilated edges detected using Canny filter

identical results obtained for masking by 7×7 and 9×9 pixels structuring element for the morphological dilation of edges for the outdoor sequence are caused by the presence of many details on the image. Therefore almost whole area of the image has been covered by the obtained mask in both cases. Exemplary images obtained using the bicubic interpolation and SR algorithm are shown in Fig. 3 (for better visualization of details only magnified fragments of images are presented).

As the preprocessed images captured by the camera typically used for mobile robot navigation purposes are further subjected to edge detection in order to detect some specific lines, corners or landmarks on the obtained binary image, the additional verification has been conducted by comparison of such processed

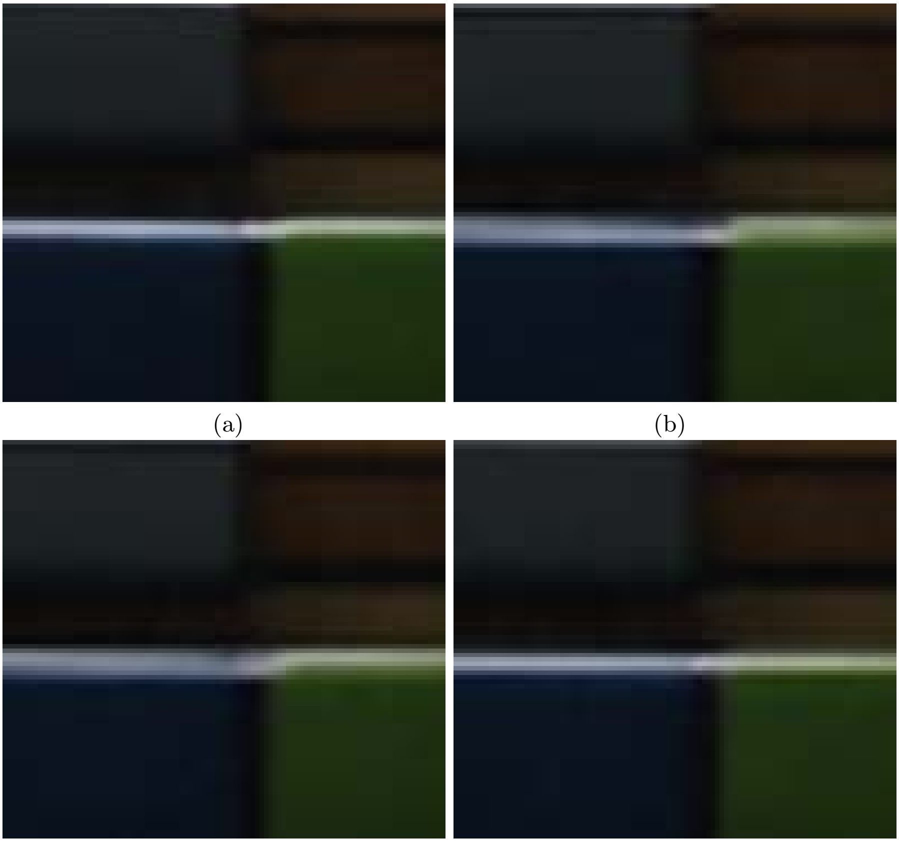
(a) (b) (c) (d)

Fig. 3. Magnified fragments of an exemplary reference image from the indoor sequence (a) and exemplary images obtained using the bicubic interpolation (b) and SR algorithm using 3 and 10 frames – (c) and (d) respectively

images, regardless of the results illustrated in Fig. 3d where the advantages of the SR algorithm are clearly visible. For this purpose the reference frames and all reconstructed images have been filtered using Canny edge detecting filter and then the comparison of output images obtained for the reconstructed frames with the result obtained for the reference image has been done. In fact, the influence of the usage of the SR algorithms on the obtained binary images may be relevant not only for the navigation of autonomous robots but also for many machine vision embedded applications, often using specialized binary image codecs [18].

Since the comparison of the binary images using the correlation coefficients or image quality (or similarity) metrics such as e.g. Structural Similarity [16] may be inadequate, some metrics typically used in binary classification have been applied. Similar metrics are typically used also for the verification of image binarization algorithms. Comparing the binary images woth the reference one,

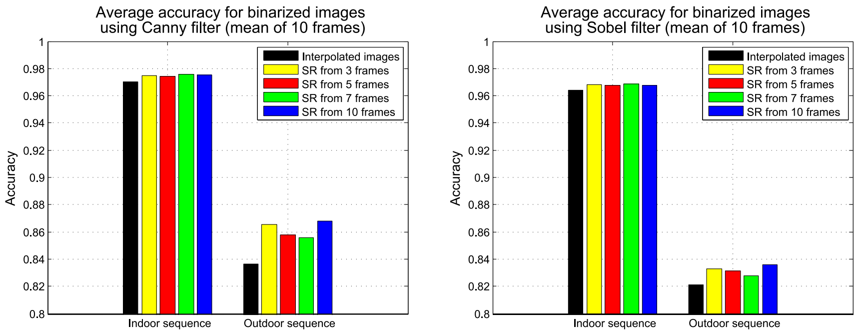

Fig. 4. Illustration of average edge detection accuracy in comparison to the reference image for reconstructed images subjected to Canny and Sobel filters

it is easy to determine the number of true positives (TP), true negatives (TN), false positives (FN) and false negatives (FN), leading to precision and recall as well as the F-Measure, specificity and accuracy and some other metrics such as e.g. Matthews correlation coefficient (MCC) [19]. Since all the metrics lead to similar conclusions, we have focused on the accuracy defined as:

$$Accuracy = \frac{TN + TP}{TP + FP + TN + FN} \ . \tag{1}$$

The experimental results obtained using Canny filter are illustrated in Fig. 4 as the average edge detection accuracy for the indoor and outdoor sequences achieved for the interpolated images and the outputs of the SR algorithm based on different numbers of frames. Higher values of the edge detection accuracy can be considered as the confirmation of the increase of the self-positioning (according to the specified patterns visible on the image) and vision based navigation accuracy of the mobile robot.

4 Conclusions and Future Work

Analyzing the results obtained for two acquired "raw" video sequences without any preprocessing operations, a strong influence of details present on the image can be easily noticed. The sequences chosen for the experiments can be considered as "hard" for the robot navigation purposes mainly due to the light conditions causing reflections of light on the floor and the presence of and many details (sett paved road, trees). Nevertheless, achieved results are promising both by means of the SSIM metric's values for the most relevant fragments of images reconstructed using the super-resolution algorithm, presented in Fig. 2, and the edge detection accuracy. The advantages of the SR algorithm can also be observed in Fig. 3 where the representation of the seat's horizontal edge is much more blurred for the bicubic interpolation (Fig. 3b) and the image obtained using the SR algorithm based only on three frames (Fig. 3c).

Due to the presence of many details in the outdoor sequence, the results of edge detection for the reconstructed frames are much worse than obtained for the indoor movie. Nevertheless, analyzing the results shown in Fig. 4 a greater relative improvement of the edge detection accuracy can be observed applying Canny filter for the outdoor video frames. Slightly worse results can be observed for Sobel filter with binary output.

Considering the assumed choice of the SR algorithm in view of its computational demand, caused by the specific area of applications, our future experiments will concentrate on the selection and possible modifications of some other image registration and reconstruction methods leading to further increase of the obtained accuracy. Further implementation of the chosen algorithms in the robotic system will allow the experimental verification of the dynamic properties of chosen methods for the real-time robot navigation.

References

1. Chatterjee, A., Rakshit, A., Singh, N.N.: Vision Based Autonomous Robot Navigation. SCI, vol. 455. Springer, Heidelberg (2013)
2. DeSouza, G.N., Kak, A.C.: Vision for mobile robot navigation: A survey. IEEE Trans. Pattern Anal. Machine Intell. 24(2), 237–267 (2002)
3. Dalgleish, F.R., Tetlow, S.W., Allwood, R.L.: Vision-based navigation of unmanned underwater vehicles: A survey. Part I: vision based cable-, pipeline- and fish tracking. Proc. Inst. Marine Engineering, Science and Technology. Part B, Journal of Marine Design and Operations B(7), 51–56 (2004)
4. Dalgleish, F.R., Tetlow, S.W., Allwood, R.L.: Vision-based navigation of unmanned underwater vehicles: A survey. Part II: vision based station keeping and positioning. Proc. Inst. Marine Engineering, Science and Technology. Part B, Journal of Marine Design and Operations B(8), 13–19 (2004)
5. Chang, C.-K., Siagian, C., Itti, L.: Mobile robot vision navigation & localization using Gist and Saliency. In: Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Systems (IROS), pp. 4147–4154 (2010)
6. Se, S., Lowe, D., Little, J.: Vision-based global localization and mapping for mobile robots. IEEE Trans. Robotics 21(3), 364–375 (2005)
7. Bonon-Font, F., Ortiz, A., Oliver, G.: Visual navigation for mobile robots: A survey. Journal of Intelligent and Robotic Systems 53(3), 263–296 (2008)
8. Vandewalle, P., Süsstrunk, S., Vetterli, M.: A frequency domain approach to registration of aliased images with application to super-resolution. EURASIP Journal on Applied Signal Processing, Article ID 71459, 14 (2006)
9. Lucchese, L., Cortelazzo, G.M.: A noise-robust frequency domain technique for estimating planar roto-translations. IEEE Trans. Signal Process. 48(6), 1769–1786 (2000)
10. Keren, D., Peleg, S., Brada, R.: Image sequence enhancement using sub-pixel displacement. In: Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), pp. 742–746 (1988)
11. Irani, M., Peleg, S.: Super resolution from image sequences. In: Proc. IEEE Int. Conf. Pattern Recognition, vol. 2, pp. 115–120 (1990)
12. Irani, M., Peleg, S.: Improving resolution by image registration. Graphical Models and Image Processing 53(3), 231–239 (1991)

13. Chatterjee, P., Mukherjee, S., Chaudhuri, S., Seetharaman, G.: Application of Papoulis-Gerchberg method in image super-resolution and inpainting. Comput. J. 52(1), 80–89 (2009)
14. Zomet, A., Rav-Acha, A., Peleg, S.: Robust super-resolution. In: Proc. Int. Conf. Computer Vision and Pattern Recognition (CVPR), vol. 1, pp. 645–650 (2001)
15. Pham, T.Q., van Vliet, L.J., Schutte, K.: Robust fusion of irregularly sampled data using adaptive normalized convolution. EURASIP Journal on Applied Signal Processing, Article ID 83268, 12 p. (2006)
16. Wang, Z., Bovik, A., Sheikh, H., Simoncelli, E.: Image quality assessment: From error measurement to Structural Similarity. IEEE Trans. Image Proc. 13(4), 600–612 (2004)
17. Li, C., Bovik, A.: Three-component weighted Structural Similarity index. In: Proc. SPIE. Image Quality and System Performance VI, vol. 7242, p. 72420Q (2009)
18. Khursheed, K., Imran, M., Ahmad, N., O'Nils, M.: Bi-level video codec for machine vision embedded applications. Elektronika Ir Elektrotechnika 19(8), 93–96 (2013)
19. Fawcett, T.: An introduction to ROC analysis. Pattern Recognition Letters 27(8), 861–874 (2006)