Real-time Tag Recognition Based on Morphology and Local Contrast

Zhiqin Chen, Yufeng Zhang, Hesheng Wang, *Senior Member, IEEE* , and Weidong Chen, *Member, IEEE*

Abstract—**With the development of computer and robotic technology, more and more miscellaneous work has been done by robots rather than men. In this case, robot vision, especially object recognition, which can help to locate objects in captured images and extract valuable information from those objects, becomes a significant part in robot design and realization. This process is expected to be real-time. In this paper, the aim is to find out the real-time and accurate algorithms to localize and recognize particular tags. Firstly, new special tags are designed. Then two tag corner localization methods are proposed, in which one is based on morphology and another is based on local contrast. And the character recognition based on OCR are also described. In the end, based on Library Management Robot (LMR), some experiments are taken. As a result, the two tag localization algorithms turn out to have complementary features in recognition distance. Meanwhile, comparing to other methods, the new methods can quickly and accurately extract tags and their information under various conditions including unexpected shadows, different light conditions and different backgrounds.**

I. INTRODUCTION

Currently, researches on service robots have been really active, which can be discovered by observing on the projects conducted by the most important conferences on robotics (IROS, ICRA, etc.), all over the world. Commonly, a mobile robot is essential to compose these kinds of systems. For example, plenty of applications have been realized, focusing on cleaning robots, rescue robots, and so on. However, to make some other applications be used in a real life scenario is rather extraordinary, like a library [1], where a manipulation is in need indeed [2]. In this last case a mobile robot arm with robust robot vision is essential to manage the books on shelves [3]. So how to find and recognize the target books is an important problem to handle. RFID technology is commonly used, which can make tag recognition fast and accurate. But it has some drawbacks such as short-distance [4], easy-interferential and difficult to handle tags simultaneously [5][6]. None of these disadvantages has been solved properly, which makes RFID unsuitable for some practical situations. As a matter of fact, tag recognition via robot vision is a better choice and there are many ways to process the captured image. [7] aims to recognize a colored tag design for Robot Soccer. Yet in library, using color to distinguish book tags from the background is unrealistic, for there are always some books which have the same color as the tags. [8] proposed a CNN (convolutional neural network) based method for tag localization and recognition. However, for the machine learning is not reliable enough so far, the experiment shows that this method is time consuming and the accuracy is not superior. [9] and [10] proposed a tag design using small pattern of black squares on a white background, and developed a morphologic approach to recognize it. It is robust when the image is clear enough, but the pattern is too large for a book spine and the algorithm is rather complicated and time-consuming. In all, there is no proper method to localize and recognize tag robustly and efficiently, especially in the library scenario. Therefore, we try to solve this problem from some different standing points.

In this paper, two real-time tag recognition methods are proposed mainly to deal with the recognition of the little-sized tags located in complex background, taking the tags on books in the library for example. So far, some preceding works about this problem have been developed in many programs, such as robot navigation strategies[9] or visually guided grasping modules[11], and so on. Specifically, this paper concentrates on the methods of localizing and recognizing tags, necessary to be real-time and accurate. The most vital question is how to precisely distinguish a tag in complicated environment, to which a reasonable solution is to localize the corners of a tag at first. Consequently, two algorithms to localize the tag corners are came up while one is based on recognizing the morphologic feature of a picture which is produced by the high pass image filter and another is based on the different local contrast of a picture, which is produced by the cartoon texture image filter. The two proposed algorithms suit for different scale of viewing distance. And according to them, several special tags are designed, which are proved to work better than the existing normal tags. After that, we describe the method based on OCR [12] to recognize the characters and information in the tags. At last, based on our Library Management Robot (LMR), some experiments about the two algorithms are taken. As a result, the two tag recognition algorithms turned out to have some different characters which are complementary and can satisfy the requirements of real-time and accuracy.

This work was supported in part by Shanghai Rising-Star Program under Grant 14QA1402500, in part by the Natural Science Foundation of China under Grant 61473191, in part by the Science and Technology Commission of Shanghai Municipality under Grant 15111104802. *(Corresponding Author: Hesheng Wang)*

Z. Chen is with Department of Computer Science, Shanghai Jiao Tong University, Shanghai 200240, China.(email: czq142857@sjtu.edu.cn)

Y. Zhang, H. Wang and W. Chen are with Department of Automation, Shanghai Jiao Tong University and Key Laboratory of System Control and Information Processing, Ministry of Education of China, Shanghai 200240, China.(email: zhang_yufeng@sjtu.edu.cn; wanghesheng@sjtu.edu.cn; wdchen@sjtu.edu.cn)

II. TAG DESIGN

For the two algorithms to localize the tags are based on features of morphology and local contrast, it is obvious that the appearance of tags influences greatly on the algorithms.

- As for the serial number tag which is attached to the spine, the current tags used in library cannot work out well because the patterns on books are easily confused with them, which also makes using color [7] to distinguish the tag from the background impossible.
- As for the two-dimensional code tag on the ground, for the camera is not absolutely top-view, the perspective view needs to be considered which may distort the tags captured by the computer.

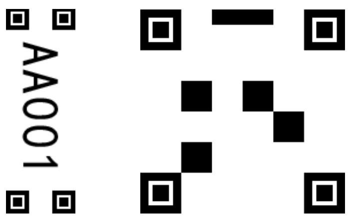
Therefore, the existing designs of tags are not suitable enough for recognizing and localizing with robot vision and then some specially designed tags come out, which are suitable for aforementioned recognition approaches.

One design is based on morphology. Four corners are added on the tag. Each tag corner has three layers, which are black center, white interlayer and black outmost layer. This specific pattern can be easily recognized, as depicted in Fig. 1.

Another design is based on local contrast. Noticing that few books have spines full of small characters or dense patterns, and floors rarely have small complicated patterns, the four corners of the tag are fitted with small dots, which can generate a high local contrast areas when the tag is captured in an image. This design is depicted in Fig. 2.

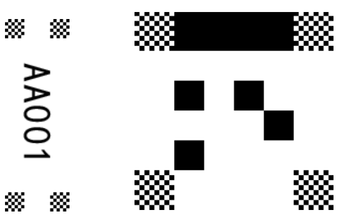
III. TAG LOCALIZATION

In this section, two methods of tag localization are developed: one is based on morphology and another is based on local contrast. The purpose is to recognize and determine the tag corners in images robustly and then localize the tag by distinguishing the unique pattern formed by tag corners. Fig.3. shows the procedure which will be described in details in the following part.



a. Serial number tag b. Two-dimensional code

Fig. 1. The first design of tags



a. Serial number tag b. Two-dimensional code

Fig. 2. The second design of tags

A. Tag corner localization based on morphology

1) High pass filter

To avoid background light or shadow affecting tag brightness, a simple high pass filter [13] is used as:

$$h_{ij} = p_{ij} - \frac{1}{(1+2\sigma)^2} \sum_{m=-\sigma}^{m\le\sigma} \sum_{n=-\sigma}^{n\le\sigma} p_{(i+m)(j+n)} \tag{1}$$

Where p represents the original image, h represents the high pass image and σ is the parameter that determines blur range. The high pass image is depicted in Fig. 5(b).

2) Extracting possible corner points

The pattern on the corners of the tag is depicted in Fig. 4 which has black center, white interlayer and black outmost layer. And an easy way to identify it is to scan the image line by line horizontally and vertically to find black pixels surrounded by white and black strips. The two red lines in Fig. 4 can demonstrate the process, while each line has crossed three black strips. So the middle strip may contain corner points, and then the two lines determine the corner point.

In order to simplify the process, we do the extraction on the binary image derived from the high pass image with a proper threshold, as depicted in Fig. 5(c). In a line of pixels, define gap g_i=1 if the two adjacent pixels p_i,p_{i+1} have different values and g_i=0 if the values are the same. When scanning a line of pixels, the number of gaps in the line can be calculated as:

$$s_i = \sum_{j=-\sigma}^{j\le\sigma} g_{i+j} \tag{2}$$

Where s_i is the number of gaps around pixel p_i, σ is the parameter that determines count range. If s_i is exactly 6, which indicates that the pixel may be a possible corner point, the pixel will be marked. When line scanning horizontally and vertically is finished, points will be checked to find whether they have been marked two times and if so, they are considered to be possible corner points.

This method can also work out well even when the tag is rotated or the image is not clear enough.

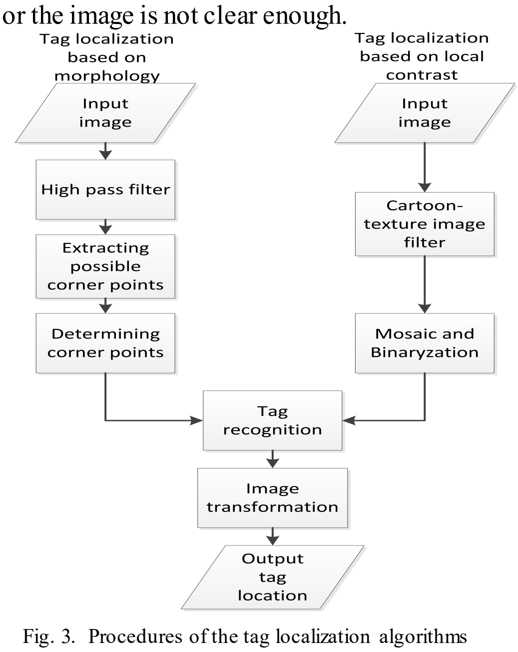


Fig. 3. Procedures of the tag localization algorithms

3) Determining corner points

After extracting possible corner points, those extracted points still cannot be determined as true corner points, for there may be other points that have the same features as corner points. So another screening called point expansion is proposed. For each possible corner pixel, assuming the pixels as an area, expand the area by adding adjacent pixels. If a pixel is adjacent to the area and the corresponding pixel in the binary image is black, add it into the area, and repeat the process until there is no new point added. This process is called black area expansion. After expanding black area, expand white area and then do black area expansion again. The number of pixels of the area can be a gauge to determine whether the start pixel is a corner pixel or not, for the area of the corner pattern is limited to a range. As a result, if the start pixel is actually a corner pixel, the whole area of the corner pattern is obtained. Expanded areas are depicted in Fig. 5(d).

If the first point is determined to be a corner point, save the weight center of the area as a true corner point, as the white points depicted in Fig. 5(a).

B. Tag corner localization based on local contrast

1) Fast Cartoon-Texture Image Filter

The Fast Cartoon-Texture Image Filter [14] can decompose an image into a geometric part and a textural part, which are also called cartoon part and texture part. A parameter, sigma, will determine the texture scale, measured in pixel mesh. Sigma has a great impact on the separation. The higher sigma is, the more cartoon part is blurred, and the more cartoon part contains.

In this algorithm, the image captured by camera is firstly processed by Cartoon-Texture Image Filter with sigma = 1, while the texture part is depicted in Fig. 6(b). Then the difference image (Fig. 6(c)) can be obtained by calculating the difference between the texture image and its background color.

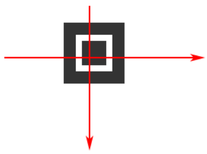
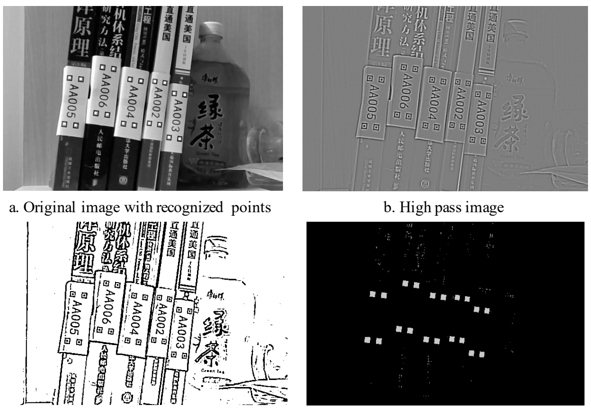


Fig. 4. The corner of the first design of tags



a. Original image with recognized points b. High pass image

c. Binary image d. Point expansion

Fig. 5. Extraction

2) Mosaic and Binaryzation

Then as the image depicted in Fig. 7(a), "mosaic" the image of Fig. 6(c) by dividing the image into 12pixel×12pixel blocks and make the block color be the average color of the block. Then make a binaryzation [15] with an appropriate threshold, where the average color of the whole image can be an option, and the processed image is depicted in Fig. 7(b), which is a map of all possible corners, and the white blocks are possibly one of the four corners of a tag.

After then, some rough positions of possible corners are obtained (Each position is the center of a white block).For each point, in order to make it more accurate, the accurate position is recalculated as:

$$\boldsymbol{P}' = \frac{\sum_{m=-\sigma}^{m\leq\sigma}\sum_{n=-\sigma}^{n\leq\sigma}\boldsymbol{P}_{(i+m)(j+n)}w_{(i+m)(j+n)}}{\sum_{m=-\sigma}^{m\leq\sigma}\sum_{n=-\sigma}^{n\leq\sigma}w_{(i+m)(j+n)}} \qquad (3)$$

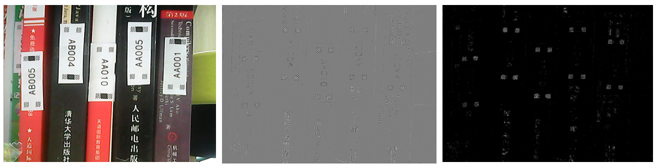
Where $\boldsymbol{P'}$ is the accurate position, (i, j) is the rough position, $\boldsymbol{P}_{ab}$ represents (a, b), w_{ab} is the gray value of pixel (a, b) of the difference image (Fig. 6(c)), and σ is the parameter that determines calculation range.

C. Tag Recognition

After obtaining the map of all possible corners, extracting all tags can be done by limiting four corners' relevant positions. Supposing the top-left point's position is (x_1, y_1), the top-right point's position is (x_2, y_2), the bottom-left point's position is (x_3, y_3), and the bottom-right point's position is (x_4, y_4), there is :

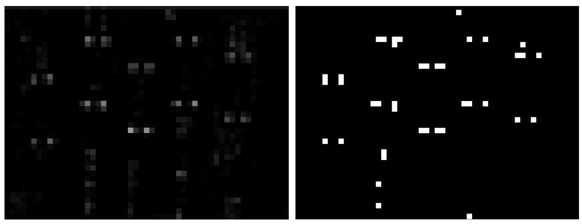
$$\begin{cases} r_{2min} \leq (x_2 - x_1) \leq r_{2max} \\ |y_2 - y_1| \leq c_{2max} \\ |x_3 - x_1| \leq r_{3max} \\ c_{3min} \leq (y_3 - y_1) \leq c_{3max} \\ |(x_4 - x_3) - (x_2 - x_1)| \leq a_{4max} \\ |(y_4 - y_2) - (y_3 - y_1)| \leq a_{4max} \end{cases} \qquad (4)$$

Where symbols like r_{2min} are constant values that constraint the coordinates. This formula is for top view images where perspective is not considered. For perspective view images, the formula is:



a. the original image b. the texture part c. the difference image

Fig. 6. Cartoon-texture decomposition



a. Mosaic image b. binaryzation image

Fig. 7. Mosaic and binaryzation image

$$\begin{cases} r_{2min} \leq (x_2 - x_1) \leq r_{2max} \\ |y_2 - y_1| \leq c_{2max} \\ |x_3 - x_1| \leq r_{3max} \\ c_{3min} \leq (y_3 - y_1) \leq c_{3max} \\ r_{4min} \leq (x_4 - x_3) \leq r_{4max} \\ |y_4 - y_3| \leq c_{4max} \end{cases} \quad (5)$$

This method is efficient and easy to modify accordingly and the outcome is depicted in Fig. 8.

What's more, with the help of four corner positions and the image transformation process, the four corner points can be accurately marked out at the four corners of a picture of our tag size, as depicted in Fig. 9.

IV. Character Recognition

This section is responsible to translate the pixels associated with images of possible tags, to authentic text tags, formed by characters, as can be found in the corresponding database, see Fig. 10.

A. Binaryzation

In order to avoid the influence of background light and shadows, Cartoon-Texture Image Filter with sigma = 3 are used to process the image and then the binary image of the texture image with an appropriate threshold is obtained, as depicted in Fig. 11.

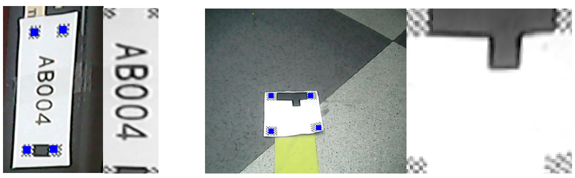
B. Partition

As depicted in Fig. 12, the process starts by scanning each column to check the pixels and if there are black pixels, mark the column as occupied (yellow part). Otherwise, mark the column to be vacant (white part). Since all Arabic numerals and English letters are continuous when projected on the bottom line, all five characters can be extracted by looking for continuous series of occupied columns (yellow blocks).

C. Matching

Now with the images of single characters, what should be done is to match them in a character library and get the numbers or letters they represent. This method can be achieved by finding the character image in the library which coincides most with the extracted character image. It's a quite efficient way and meanwhile most of the characters can be translated correctly.

Fig. 8. Corner points of possible tags



a. Serial number tag b. Two-dimensional code

Fig. 9. Transformation result

V. Experiments

A. Experiments based on LMR Platform

The experiments are taken on our Library Management Robot (LMR), which is designed to autonomously find and grasp books in library with the help of robot vision.

As shown in Fig. 13(a), the experiments can be done with several modules of the LMR. The grasping module is made up of the telescopic pole and special fingers for picking up books. And the camera, Logitech C930e with the resolution of 1920*1080, that provides vision for the LMR to localize and recognize books is set on this module. The distance between the camera and books with tags will change when the telescopic pole works, see Fig. 13(b).

1)Comparing experiment on tag recognition methods

As a mainly concerned problem is the influence of distance between camera and tags on the recognition methods, an experiment are designed to change the distance for 2cm every step and in each step and the methods are testified by 100 books with tag.

It's estimated that the overall Tag Localization Rate (TLR) is greatly influenced by the distance, which is shown in the Fig. 14. Two methods both show great accuracy in their own suitable distance where the TLR is almost 100%, but if out of the suitable distance, the TLR drops very quickly.

As a result, the two methods turn out to have the complementary features when distance change, while the method based on morphology performs well in medium and long distance (34-70cm) and meanwhile, the method based on local contrast works better in short distance (13-40cm).

It's realized that the TLR mostly depends on the Tag Corners Localization Rate (TCLR) and Tag Misrecognition Rate (TMR) of the two tag corner localization algorithms. Therefore, some analyses are done based on the experiment data, to find reasons for the aforementioned conclusions.

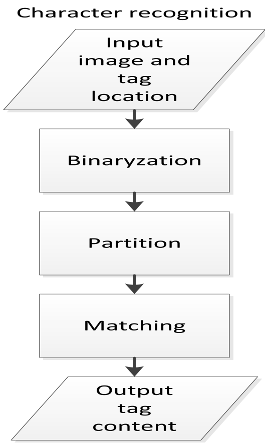


Fig. 10. Procedures of the character recognition

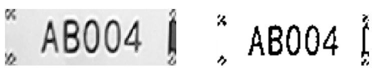


Fig. 11. Binaryzation result

Fig. 12. Partition

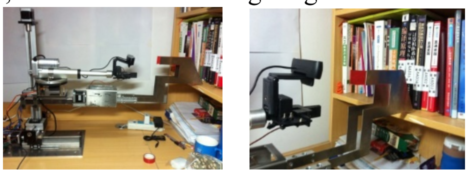
2)Analysis: tag corner localization based on morphology

The minimum distance is around 34cm. Fig.15 (a) and (b) shows the reason that for the size of tags has been very large and if getting closer, the TCLR drops dramatically.

As the distance increases, the tags corners can be located robustly while the TMR also goes down and when the distance comes to 45cm, the misrecognized points appear. As Fig.15 (b) shows, although the top two points are the other characters which are misrecognized, for the reason that it needs four corner points which can form a rectangle to represent a tag, the tag recognition is not impacted by the misrecognition.

The maximum distance turns out to be approximate 71cm. From Fig.15 (a), in this distance, the TCLR drops to 85% and at least 5% characters are misrecognized. These two factors begin to worsen the TLR.

This algorithm is greatly influenced by the distance, because it is based on some morphologic features of the tag corner, in which the size of legal tag corners are restricted within an interval. It means only the objects in this interval can have the possibility to be recognized as a tag corner and this threshold makes a positive effect on the accuracy of the tag corner localization, which reduces the risk of interruption from other objects. However, if the camera is too close to the tags, which the tag corners become too big in the picture, the success rate will reduce dramatically. One solution is to modify the threshold according to the distance.



a.Overall platform b. The grasping module with camera

Fig. 13. Details about the camera

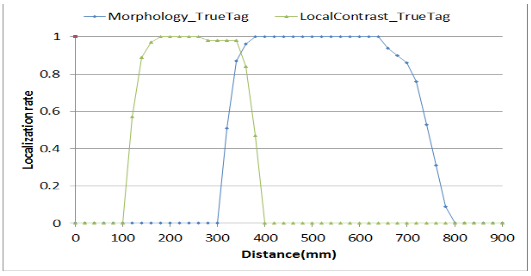
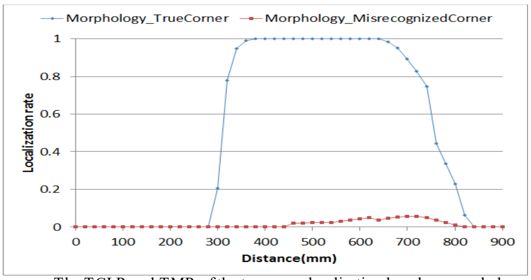
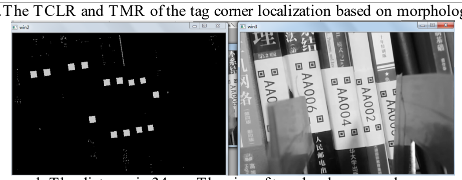


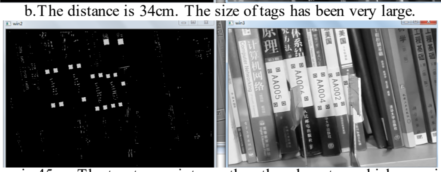
Fig. 14. The overall Tag Localization Rate (TLR) of two methods



a.The TCLR and TMR of the tag corner localization based on morphology



b.The distance is 34cm. The size of tags has been very large.



c.The distance is 45cm.The top two points are the other characters which are misrecognized.

Fig. 15. Experiment on tag corner localization based on morphology

3)Analysis: tag corner localization based on local contrast

It is known that when the distance is above 40cm, the tag can hardly be recognized and Fig. 16(a) can explain this. In that distance, the TCLR falls down under 60% and the TMR keeps at 10% or more, since the main cause is the restriction of camera's resolution. The designed tag corner based on local contrast is complexly made up by dozens of white and black little blocks. When the resolution is not enough, these corners lost their characters and can't be clearly recognized by camera.

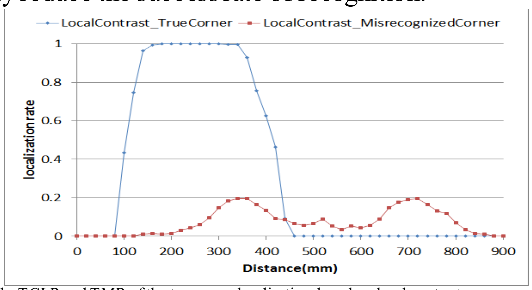
However, as shown in Fig. 16(a), the tag corners can be robustly identified even if the distance is just around 13cm and with a 5% TMR. It is because the local contrast of corners is always higher than others', that the tags can be localized in an extremely short distance, which is shown in Fig. 16(b). And this feature provides a possibility that the tag recognition method can be used in a larger distance scale than others.

Comparing experiments with related works

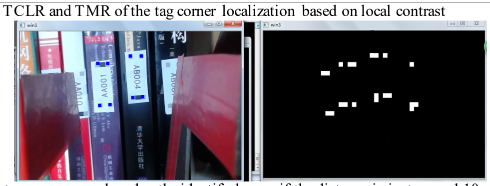
1) Method from the UJI librarian robot

This experiment is designed to compare our approaches with the method of the UJI librarian robot [3]. Various conditions have been considered, such as brightness, shadows, colors and aging tags. As a result, it shows that UTJ's method has several flaws, which using our corner localization based method can effectively avoid.

According to Fig.17, although the UJI's method can adapt different light, it is extremely sensitive to shadows and brightness gradual change, since it processes the image only by Histogram Equalization and Binarization, even while in normal conditions, some tags may not be recognized because of shadow or influenced by other patterns (e.g. book names). Meanwhile, aging tags can be also confusing, which shows that what makes the tag less bright than the background can greatly reduce the success rate of recognition.



a.The TCLR and TMR of the tag corner localization based on local contrast



b.The tag corners can be robustly identified even if the distance is just around 10cm

Fig. 16. Experiment on tag corner localization based on local contrast

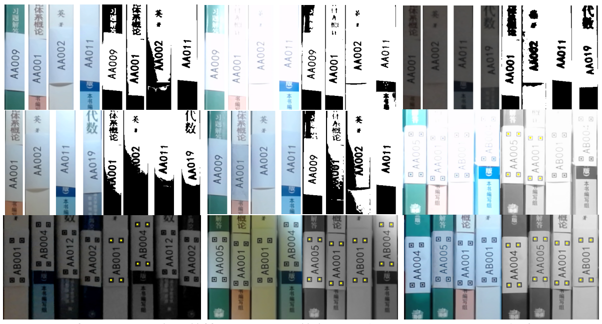


Fig. 17. Performance in different conditions. *(In every image, left part is the origin image and the other is the processed image. From upper left to lower right, the first five images show the performance of the UJI's method in normal, very bright, very dim, shadow and aging tag (AA001) conditions accordingly. Although individual books are divided, the result of recognition is not robust. And the last four images show the performance of morphology based method (their corners are marked yellow) in very bright, very dim, shadow and aging tag (AA005) conditions, which is in high success rate.)*

However, with the help of the morphology based corner localization algorithm and the filter, tags can be recognized in high success rate in spite of reasonable brightness changes and shadows. In addition, the UJI's method uses the same character partition technique with ours, but there is not image transformation. If a book is placed askew, misrecognition of the characters will appear because the image is distorted and the projection of characters becomes difficult to distinguish since projections of single characters are connected together.

2) Machine learning

During the research, many related works based on machine learning have been searched, but none of them can work out well for this problem. It can be inferred that the main difficulty is that machine learning methods search specific features all over the image with high time-complexity, which does not meet the requirement of real-time easily.

B. Other preliminary results

Some other experiments on the speed and accuracy have also been done. Here are some preliminary results:

1. The whole tag recognition process, including tag localization and characters recognition, is contended with the need of real-time. Around 25 pictures can be processed every second and the delay is minor.

2. The character recognition based on OCR is accurate enough. While the camera is within 70cm from the tags, the recognition rate is about 93% to 97% in a single picture.

VI. Conclusion

In this paper, real-time and accurate methods to localize and recognize particular tags are presented. Firstly, some new tags were designed. Then two tag corner localization methods were proposed, in which one is based on morphology and another is based on local contrast. After that, we described the character recognition method based on OCR. In the end, some experiments and comparisons were taken. As a result, the two tag localization algorithms turned out to have some complementary characters in recognition distance. And the new methods can quickly and accurately extract tags and their information under various conditions including unexpected shadows, different light conditions and different backgrounds.

References

[1] Suthakorn, J., et al. "A robotic library system for an off-site shelving facility." *Robotics and Automation, 2002. Proceedings. ICRA '02. IEEE International Conference on* IEEE, 2002:3589-3594 vol.4.

[2] Kim, Bong Keun, et al. "Design and control of the librarian robot system in the ubiquitous robot technology space." *Robot and Human Interactive Communication, 2008. RO-MAN 2008. The 17th IEEE International Symposium on* IEEE, 2008:616 - 621.

[3] Prats, Mario, et al. "The UJI librarian robot." *Intelligent Service Robotics* 1.4(2008):321-335.

[4] Flores, Jose Luis Martinez, et al. "Performance of RFID tags in near and far field." *Personal Wireless Communications, 2005. ICPWC 2005. 2005 IEEE International Conference on* IEEE, 2005:353-357.

[5] Hwang, Tae Wook, et al. "An Anti-collision Method for Fast Recognition of Multiple RFID Tags." *Optical Internet and Next Generation Network, 2006. COIN-NGNCON 2006. The Joint International Conference on* IEEE, 2006:287-290.

[6] Jo, M., and H. Y. Youn. "Intelligent recognition of RFID tag position." *Electronics Letters* 44.4(2008):308-309.

[7] Yu, Qiushui, et al. "Colour tag design of robot soccer based on computational verb theory." *Electronic Measurement & Instruments, 2009. ICEMI '09. 9th International Conference on* IEEE, 2009:4-433-4-437.

[8] Bulan, Orhan, et al. "USDOT number localization and recognition from vehicle side-view NIR images." *2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* IEEE Computer Society, 2015:91-96.

[9] Matas, J., L. M. Soh, and J. Kittler. "Object recognition using a tag." *Image Processing, International Conference on* IEEE Computer Society, 1997:877-880 vol.1.

[10] Soh, L. M., J. Matas, and J. Kittler. "Model Acquisition And Matching In Tagged Object Recognition (tor)." *Signal Processing Conference (EUSIPCO 1998), 9th European* IEEE, 1998.

[11] Sanz, P. J., et al. "Vision-Guided Grasping Of Unknown Objects For Service Robots." *Robotics and Automation, 1998. Proceedings. 1998 IEEE International Conference on* 1998:3018 - 3025.

[12] *Optical Character Recognition (OCR) – How it works.* Nicomsoft.com. 2013.

[13] Paul M. *Mather. Computer processing of remotely sensed images: an introduction (3rd ed.).* p. 181. 2004.

[14] Antoni, Buades, et al. "Fast cartoon + texture image filters." *IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society* 19.8(2010):1978-1986.

[15] Milyaev S, Barinova O, Novikova T, Lempitsky V, and Kohli P, "Image binarization for end-to-end text understanding in natural images." in *International Conference on Document Analysis and Recognition (ICDAR)*,