An Object Recognition Approach based on Structural Feature for Cluttered Indoor Scene

Wenbo Yuan, Zhiqiang Cao, Peng Zhao, Min Tan
State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, CAS
Beijing, China
{wenbo.yuan, zhiqiang.cao, peng.zhao, min.tan}@ia.ac.cn

Yuequan Yang
College of Information Engineering, Yangzhou University,
Yangzhou, China
yangyq@yzu.edu.cn

***Abstract*—In this paper, aiming at the indoor scene under monitoring by visual sensor network (VSN), an object recognition approach based on structural feature is presented. Firstly, we regard the output of existing line segment detector LSD with proper parameters as the preliminary extraction result and it still will be further restored and split. Then, we give an inference model based on structural features of object including line segment ontology characteristics and relative relationship between the line segments. Finally, the objects are recognized with position information through inference. The effectiveness of the approach is verified, and the results show that our approach does not rely on segmentation and has robustness on partial defect and structural deformation to some extent.**

Keywords*—*indoor scene; object recognition; visual sensor network; structural feature; inference model

I. Introduction

With the rapid development of communication technology, the implementation of the visual sensor network (VSN) shows a trend of diversification and simplification by using Ethernet, Internet, 3G network, etc. Networks of visual sensors have recently emerged as a new type of sensor-based intelligent system that brings performance and complexity challenges[1].

As the function of VSN is more and more powerful, VSN has been widely used in areas such as computer vision, robotics, etc. Major research issues of VSNs include camera coverage optimization, network architecture, and low-power visual processing and communication, etc[2]. Czarlinska and Kundur realize the reliable event-detection by using untethered camera nodes and scalar sensors, which are deployed in a hostile environment[3]. A framework of hierarchical feature distribution for object matching in a network of visual sensors is proposed in [4]. Without prior knowledge of the target in wireless VSN, [5] proposes a sensor selection solution to improve the accuracy of target localization. A multi-view visual-target-surveillance system was built in [6], which can autonomously implement target classification and tracking with collaborative online learning and localization. [7] highlights the efficient use of the geometric constraints to derive distributed algorithms for target detection, tracking and recognition. Yokoya *et al.* propose a calibration method of a distributed vision network and a group of mobile robots are used to obtain a set of calibration data for each camera[8]. [9] presents a human behavior recognition method, which uses a tangible network robot system composed of a mobile robot and vision sensors embedded in an environment.

In order to better take advantage of the information provided by VSN, scene understanding should be focused, where object recognition plays an important role. There are two kinds of approaches to realize object recognition: image data-based and model-based. The former firstly obtain the outline or partial parts of the object based on image segmentation, then extract image features and describe the object (partial parts) by using the descriptors such as SIFT[10], shape context[11], HOG[12], etc, and finally, locate and recognize the object based on the relationship among the candidates that conform to the features of object. Although these methods have achieved a good detection effect, they rely on segmentation with complicated algorithm structure. Meanwhile the target is subject to be interfered by cluttered background, partial defect, structural deformation as well as the variance of illumination. The model-based recognition traverses the image directly to a recognition model that is built from a set of assumptions and expectation properties generated by high-level knowledge.

Extracting the environment information into line segments[13] has received attentions due to its better description about the geometric content of images, which makes the object recognition on the basis of line segments be promising. [14] proposes a linear-time line segment detector LSD that gives accurate results and a controlled number of false detections. In this paper, aiming at the indoor scene under monitoring by visual sensor network (the schematic diagram is shown in Fig. 1), we regard the output of existing LSD with proper parameters as the preliminary extraction result and it still will be further restored and split, which form line segment generating module. Then, an inference model based on structural features of object is proposed to find the line segments describing the object.

The remainder of this paper is organized as follows. Section II gives the line segment generating module. Section III presents the inference model based on structural features of object. The experiment results are described in section IV and section V concludes the paper.

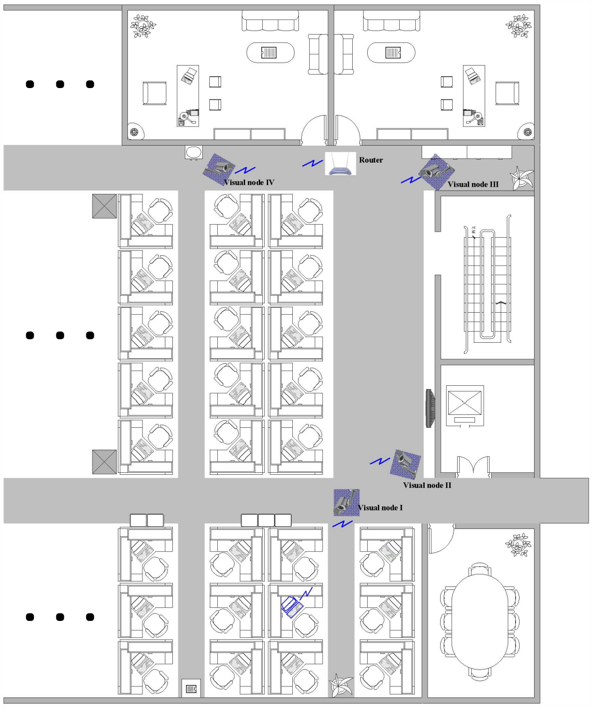


Fig. 1. The schematic diagram of an office scene under monitoring by visual sensor network, which consists of visual nodes I, II, III, IV, and a router

II. LINE SEGMENT GENERATING MODULE

In regard to the ordinary object recognition, compared with color, texture, image transformation and other features, structural feature can achieve better adaptability in changing environments. Based on structural features of object, the first task for object recognition is to extract the line segments with high accuracy, which constitute a line segment set Φ_L.

A. Line segments preliminary extraction based on LSD

LSD[14] is a line segment detector with a high accuracy and a satisfactory result. In this paper, we adopt LSD to obtain the preliminary line segments result. Considering that most object structures in indoor scene are composed of straight lines, the parameters in LSD is adjusted for better environmental adaptability. In addition, some similar line segments are simplified for reducing computation amount.

B. Line segments restoration

In a cluttered indoor scene, the object is often blocked by other objects to some degree, which will greatly interfere with structural features. In addition, there exists partial defect, structural deformation and the limitation of image quality, which may occur problems, such as the extracted line segments would be disconnected. Therefore, it is necessary to find out and repair the unconnected line segments to better reflect the related structure features.

For any two line segments A and B, if they have similar direction, small vertical distance, and the distance between their endpoints is within a certain range, they are likely to belong to one line segment. The specific criterion of judgment is as follows:

$$\begin{cases} \alpha_{BA} \le \alpha_T \\ d_{ss} \le d_T \\ d_{BA} \le (\alpha_T - \alpha_{BA}) \times a_\alpha + (d_T - d_{ss}) \times a_d + B_T \times a_B \end{cases} \quad (1)$$

where α_{BA} is the angle difference of these two line segments, d_{SS} is the minimum distance from the middle point of shorter line segment to the longer line segment, d_{BA} is the minimum distance between the endpoints of line segments, α_T and d_T are given thresholds, B_T is a Boolean variable and it is 1 if there is a third line segment between the two line segments, a_α, a_d and a_B are weighting coefficients.

If all the conditions in formula (1) are satisfied, the two line segments will be repaired into one line segment with farthest two endpoints being kept.

C. Line segments split

In some cases, a line segment belonging to the object may be extended due to complexity of the environment, which will affect the recognition effect of the object. If any two line segments with enough lengths intersect or present a T-shaped structure, the line segments need to be split and all sub-segments are saved as new line segments. Note that the original line segments will be kept as well.

III. INFERENCE MODEL BASED ON STRUCTURAL FEATURES

A. Structural features

Under certain scale change of camera view, relative relationship among the line segments is stable. Herein, two categories of structural features are chosen and they are line segment ontology characteristics and relative relationship between the line segments, as shown in Fig. 2.

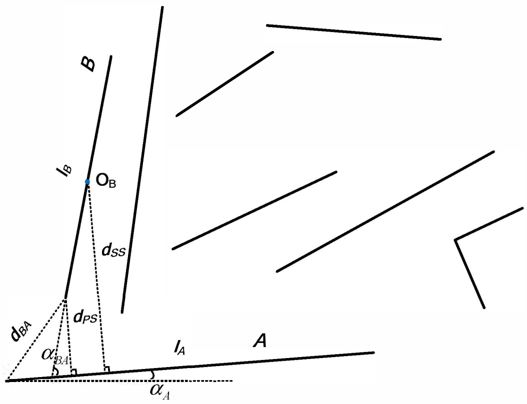


Fig. 2. Structural features of line segments A and B

The line segment ontology characteristics include its angle, length as well as spatial position in image. Take line segment A as an example, and its angle and length are α_A and l_A, respectively. Relative relationship between the line segments is given as follows. Take line segment B relative to A as an example. Besides α_{BA}, d_{SS} and d_{BA}, the following features are considered and they are length ratio R_{BA}, verticality V_{BA}, parallelism P_{BA} between the line segments, shortest distance d_{PS} between the endpoints of B to line segment A as well as relative bearing to A, where $R_{BA}=l_B/l_A$, $V_{BA}=2\alpha_{BA}/\pi$, $P_{BA}=180\alpha_{BA}/\pi$, l_A and l_B are the lengths of line segment A and B, respectively.

B. Inference model

On the basis of structural features mentioned above, the evaluation for each line segment is as follows:

$$Q = \sum_{i=1}^{m}(T_Q - F_i \times W_{F_i}) \Big/ m \tag{2}$$

where m is the number of structural features and it is determined by the task and the specific object, T_Q is the given value, F_i is the difference between the actual value and standard value S_i corresponding to the feature i, W_{Fi} is the weight corresponding to F_i. Note that the standard values vary with the poses of visual nodes and the objects to be recognized.

Assume that the object to be recognized may be described by M line segments. Firstly, the line segments in Φ_L with lower evaluation values are eliminated based on line segment ontology characteristics and the given evaluation threshold value T_B, and each passed line segment is regarded as a candidate of the first line segment of the object. For each candidate of the first line segment, based on actual relationship of M line segments and the selected structural features, the candidates with maximum evaluation values of other M-1 line segments of the object are acquired successively from Φ_L. Thus, a series of line segments groups whose amount is equal to that of the candidates of the first line segment are obtained and each group with M line segments is considered as an imaginary object. Synthesize the evaluation value of each line segment in each group, and thus the effective groups may be confirmed as the objects based on the value T_B. The detailed algorithm is shown in Fig. 3, where for line segment $L_j \in \Phi_L$, $Q_{jk}(T_Q, S_i, W_{Fi})$ is the value of L_j corresponding to k^{th} line segment of object; G is a temporary set of M line segments, V is a temporary evaluation values set corresponding to M line segments.

input: line segment set Φ_L, number M of line segments of object, T_Q, T_B, S_i, W_{Fi}.
output: objects set Φ_o, number of objects N.

$\Phi_o \leftarrow \mathbf{0}$;
$N \leftarrow 0$;
$G \leftarrow \mathbf{0}$;
$V \leftarrow \mathbf{0}$;
for each line segment in Φ_L **do**
  for $(Q_{j1}(T_Q, S_i, W_{Fi}) \geq T_B)$ **do**
    $G(1) \leftarrow L_j$;
    $V(1) \leftarrow Q_{j1}(T_Q, S_i, W_{Fi})$;
    for $(k \geq 2$ and $k \leq M)$ **do**
      $t^* \leftarrow \arg\max_{L_t \in \Phi_L}(Q_{tk}(T_Q, S_i, W_{Fi}))$;
      $G(k) \leftarrow L_{t^*}$;
      $V(k) \leftarrow Q_{t^*k}(T_Q, S_i, W_{Fi})$;
    end
    if $(sum(V)/M \geq T_B)$ **then**
      if $(match(G, \Phi_o))$ **then**
        the better one is kept;
      else
        add G to Φ_o;
        $N \leftarrow N+1$;
      end
    end
  end
end

Fig. 3. Algorithm of inference for object recognition

IV. EXPERIMENTAL RESULTS

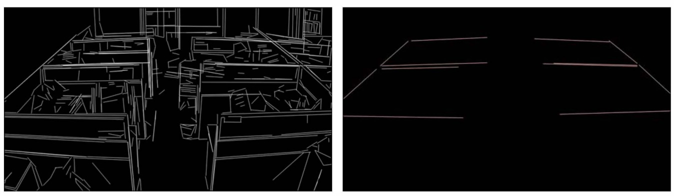
In the following experiments, our visual sensor network is built in an office scene, which consists of four visual nodes and a router. Visual nodes are installed in the office with appropriate postures and they are networked in an IP way. All images are transferred to the upper platform via wireless communication.

The parameters[15] of existing LSD in line segments preliminary extraction are chosen as follows: S=0.6, Σ=0.45, $\rho=2\sqrt{2}$, τ=45, ε=1, D=0.5. The other parameters are given by $\alpha_T=2\pi/90$, d_T=3, a_α=10, a_d=10, a_B=15, T_Q=10, T_B=9.

This paper mainly focuses on the recognition of objects in the cluttered indoor scene, such as cubicle, fire hydrant service container and escape route sign.

The cubicle is considered to be formed by three line segments that present a U-shaped structure with conspicuous vertical or parallel relationship, which provide the basis of structural features. Fig. 4 gives the process of the structural feature-based recognition approach with the identification of the cubicles. For the image depicted in Fig. 4(a) captured by visual node I, Fig. 4(b) and Fig. 4(c) give the outputs of existing LSD and line segment generating module, respectively. The recognition results are shown in Fig. 4(d). From Fig. 4(e) with the recognition results shown in the original image, it is seen that the cubicles are recognized with better matching. Furthermore, the positions of cubicles may be obtained due to known position information of the outputted line segments. Fig. 5(a) and Fig. 5(b) give the recognition results of cubicles corresponding to original images captured by visual node II and visual node III, respectively.

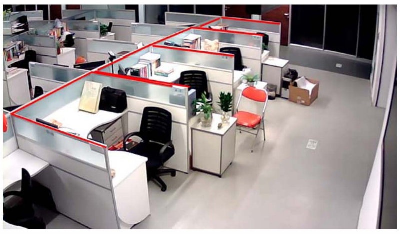
(a) Original image (b) The output of existing LSD



(c) The output of line segment generating module (d) The recognition results

(e) The recognition results shown in the original image

Fig. 4. The cubicles recognition process of an original image captured by visual node I



(a) Visual node II

(b) Visual node III

Fig. 5. The recognition results of cubicles shown in the original images

For the fire hydrant service container, it consists of eight line segments, which present two cascaded rectangles. In addition, the escape route sign is a rectangle with four line segments. Fig. 6 gives the recognition results corresponding to original image captured by visual node IV. It is seen that the fire hydrant service container and the escape route sign are recognized effectively although the former has partial defect caused by a cubicle.

Fig. 6. The recognition results of fire hydrant service container and the escape route sign shown in the original image captured by visual node IV

V. CONCLUSION

Aiming at the indoor scene under monitoring by visual sensor network, we present an object recognition approach based on structural feature. From the experiments we have conducted, the approach achieves a better recognition rate with position information. In the future, we plan to improve the performance of line segments restoration and split to obtain more accurate line segment information for the inference model. In addition, integrating the recognition results of multiple images from visual nodes into one topological map will be considered.

ACKNOWLEDGMENT

This work was supported in part by the National Natural Science Foundation of China under Grants 61273352, 61175111, 61227804, 60805038.

REFERENCES

[1] A. Marcus, and O. Marques, "An eye on visual sensor networks," IEEE Potentials, vol. 31, no. 2, 2012, pp. 38-43.

[2] Y. Charfi, N. Wakamiya, and M. Murata, "Challenging issues in visual sensor networks," IEEE Wireless Communications, vol. 16, no. 2, 2009, pp. 44-49.

[3] A. Czarlinska, and D. Kundur, "Reliable event-detection in wireless visual sensor networks through scalar collaboration and game-theoretic consideration," IEEE Transactions on Multimedia, vol. 10, no. 5, 2008, pp. 675-690.

[4] V. Sulic, J. Pers, M. Kristan, and S. Kovacic, "Efficient feature distribution for object matching in visual-sensor networks," IEEE Transactions on Circuits and Systems for Video Technology, vol. 21, no. 7, 2011, pp. 903-916.

[5] W. Li, and W. Zhang, "Sensor selection for improving accuracy of target localisation in wireless visual sensor networks," IET Wireless Sensor Systems, vol. 2, no. 4, 2012, pp. 293-301.

[6] X. Wang, S. Wang, and D. Bi, "Distributed visual-target-surveillance system in wireless sensor networks," IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, vol. 39, no. 5, 2009, pp. 1134-1146.

[7] A.C. Sankaranarayanan, A. Veeraraghavan, and R. Chellappa, "Object detection, tracking and recognition for multiple smart cameras," Proceedings of the IEEE, vol. 96, no. 10, 2008, pp. 1606-1624.

[8] T. Yokoya, T. Hasegawa, and R. Kurazume, "Calibration of distributed vision network in unified coordinate system by mobile robots," Proceedings of 2008 IEEE International Conference on Robotics and Automation, Pasadena, United states, 2008, pp. 1412-1417.

[9] K. Kemmotsu, Y. Koketsu and M. Iehara, "Human behavior recognition using unconscious cameras and a visible robot in a network robot system," Robotics and Autonomous Systems, vol. 56, no. 10, 2008, pp. 857-864.

[10] D.G. Lowe, "Object recognition from local scale-invariant features," Proceedings of the 7th IEEE International Conference on Computer Vision, Kerkyra, Greece, 1999, pp. 1150-1157.

[11] S. Belongie, J. Malik and J. Puzicha, "Shape matching and object recognition using shape contexts," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 4, 2002, pp. 509-522.

[12] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," Proceedings of 2005 IEEE Conference on Computer Vision and Pattern Recognition, San Diego, United states, 2005, pp. 886-893.

[13] Y. Chung, C. Wang, V.K. Prasanna, "Parallel Algorithms for Perceptual Grouping on Distributed Memory Machines," Journal of Parallel and Distributed Computing, vol. 50, no. 1-2, 1998, pp. 123-143.

[14] R.G.v. Gioi, J. Jakubowicz, J. Morel and G. Randall, "LSD: A fast line segment detector with a false detection control," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, no. 4, 2010, pp. 722-732.

[15] R.G.v. Gioi, J. Jakubowicz, J. Morel, and G. Randall, "LSD: a Line Segment Detector," Image Processing On Line, vol. 2012 [http://dx.doi.org/10.5201/ipol.2012.gjmr-lsd].