Robot localization based on visual odometry

Dan Wu[1, a]

[1]School of Electronics and Information Engineering,Beijing Jiaotong University, Beijing 100044, China

[a]wudanhxh@163.com

Keywords: Localization, Viusal Odometry, Feature, Direct,Motion Estimation.

Abstract. With the development of computer vision technology, the positioning technology based on vision sensor had drawn more and more attentions. There were two kinds of visual odometry technologies through the image information for motion estimation to obtain the attitude information, one of them was based on the feature points and the other one was the direct method without the feature points .In recent years,many scholars approached various of methods and visual odometry algorithm based on the image data but no one is perfect.KinectV1, as a high-performance RGB-D sensor, could capture both color and depth images. The evaluation about two kinds of visual odometry technologies based on KinectV1 sensor was carried out.A summary and analysis for the robustness and accuracy problem was studied and researched. The results of evaluation showed that the method based on feature points could be applied to the environment riched in features,and the direct method is more robust in the environment of visual feature degradation.

1.Introduction

With the development of robot technology, the intelligent mobile robot with moving, environment perception and planning had become the main direction of robot development. Intelligent robot as the highest level of automation system, the research and application of that played an important role in promoting the development of science and technology， improving labor productivity, and was one of the major hallmarks of the level of a national manufacturing level.

The earliest robot localization method was based on some marker that robots could identify in the environment, such as scene recognition method [1],[2] triangulation. In the same time, the laser, sonar, Radio Frequency Identification and other types of sensors were applied to solve the mobile robot localization problem [3][4] gradually.

The location[5],[6] technology based on visual odometry was one of location technology based on robot vision.Its principle was through calculating the characteristics of the image captured by the camera mounted on the robot and to realize location.Huang proposed Fast Odometry Vision based on visual odometry(FOVIS).The dense RGB-D visual odometry method proposed in the paper[7] which could reduce the photometric error between two RGB frame. Pomerleau[8] put foward a iterative closest points (ICP) visual odometry method by the depth information, although this method can run on embedded computer, but it cost too much computing resource.

The basic principle of visual odometry was according to the different environmental information acquired by robots.The environmental information was stored in an image sequence or a video stream,and the information was applied to calculate or forecast the location of the robot.

Multiple types of sensors could be used to calculate the trajectory of robot based on visual odometry such as a single camera, stereo camera and RGB-D camera. The skeleton[9] of visual odometry based on monocular and stereo vision was simple and low cost, but it could not obtain the relative distance between the objects in the image. But the RGB-D camera can directly measure the distance between each pixel in the image and the camera through the infrared light or TOF principle,so it could provide more information than traditional cameras.To sum up,many visual odometry based on RGB-D sensor [10][11] had good performance in robot localization.Currently the RGB-D camera includes Kinect, Xtion,etc.. In this paper,Kinect V1 was adopted as a data sensor.

At present, the main methods of visual odometry divided into two kinds，one was based on feature points and the other one was direct method without feature points. The method using feature point was also called the sparse method， and the direct method was called dense method.

2.Visual odometry estimation based on the feature of RGB-D

The method based on feature points was the mainstream method of VO. In this method, first,the scene images were collected. Second, some representative points called feature points would be selected. And then robot location would be realized through these feature points and the remaining would be abandoned. The procedure of the algorithm shown in figure 1.

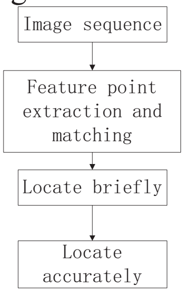


Fig.1 RGB-D vision odometry estimation algorithm based on feature method

2.1 Feature extraction

In computer vision,especially in object identification,there are many identification methods based on feature detection and extraction of object surface. In order to track these feature points,the selected features points would be invariance and stable while the robot or camera moving and rotating, but also need to have a low time complexity.There were various method of feature extraction, such as Scale Invariant Feature Transform(SIFT),Speed Up Robust Feature(SURF),corner detection (Harris) or features from accelerated segment test(FAST).

2.2 Feature matching

Feature point matching aimed to find out the corresponding relationship between the pixels of two images, so as to determine the location of the two images. Visual odometry could be regard as matching the new image with the previous frame to estimate its motion simply. There were various method of feature matching, such as feature points extraction algorithm with corresponding feature matching method(SIFT and SURF),Fast library for approximate nearest neighbors(FLANN) and Brute force method.

2.3 Motion estimation

The set of the feature got by feature extraction and matching between two consecutive frames would be collected,they were two coordinates of the three-dimensional coordinates.Motion parameters which was motion estimation would be calculated through these set.The motion parameters are represented by the rigid body transformation,the R was rotation matrix and the t was translation vector.The relationship between the motion parameters and the coordinates of the feature points was shown in formula (1):

$$P_{cj} = RP_{pj} + t \tag{1}$$

In the formula, P_{cj} and P_{pj} represented the 3D coordinates of the current frame and the corresponding J feature points of the previous frame. If there were N pair of matching points, then j=1,2,3...N.

Solving the motion parameters need 4 sets of non collinear matching points without considering the effects of camera distortion.In reality,we must consider the impact of camera calibration and the wrong feature point of the match, so the number of the feature point matched was much greater than 4.

Some false might be caused by FLANN matching,the result of motion estimation calculated by that would be not accuracy.So in order to eliminate false or errors,the algorithm of RANSAC[12]was adopt before the motion estimation.

The rotation matrix and translation vector of two consecutive frame motion transformations are calculated by experiment:

$$R=\begin{pmatrix} 1 & 0.000243628 & -0.000761814 \\ -0.0000244027 & 1 & -0.000523273 \\ 0.000761686 & 0.000523458 & 1 \end{pmatrix} \tag{2}$$

$$t=\left(0.00368371 \quad -0.00161102 \quad -0.000623746\right)^{T} \tag{3}$$

3. Visual odometry estimation based on the direct method

The direct method[13]provided a new way to locate. In the feature based method, a lot of useful information in the image were lost during the process of extracting the feature points. Because of the defects in the structure of the feature points method, it could not guarantee that the collected images could provide enough and effective feature points. Therefore, the direct method no longer extracted the feature, using all the pixels information by minimizing the gray pixel difference between consecutive frames to achieve location. The direct method was based on the assumption that the gray level of the pixel of in the same space was invariant.

$\mathrm{P}_i=(\mathrm{X},\mathrm{Y},\mathrm{Z})$ was the 3D point scene in space camera. The camera obtained two continuous frames respectively at the two moments. In the coordinate system, the second frame was the rotation and translation of the image relative to the first frame image. So the point moved and rotated as the rigid motion $\mathrm{g}\in\mathrm{SE}(3)$. The P_i pixel coordinates in two frames are p_i and $\mathrm{p'}_i$ respectively.

The spatial point and the homogeneous coordinates of the pixel points satisfied the projection relation:

$$\mathrm{p}_i=\begin{pmatrix} u_i \\ v_i \\ 1 \end{pmatrix}=\frac{1}{Z_i}K\begin{pmatrix} \mathrm{X} \\ \mathrm{Y} \\ \mathrm{Z} \\ 1 \end{pmatrix}=\frac{1}{Z_i}K\mathrm{P}_i \tag{4}$$

$$\mathrm{p}_i\,'=\begin{pmatrix} u_i\,' \\ v_i\,' \\ 1 \end{pmatrix}=\frac{1}{Z_i\,'}K\begin{pmatrix} R & t \\ 0 & 1 \end{pmatrix}\begin{pmatrix} \mathrm{X} \\ \mathrm{Y} \\ \mathrm{Z} \\ 1 \end{pmatrix}=\frac{1}{Z_i\,'}K(\mathrm{RP}_i+\mathrm{t}) \tag{5}$$

Where $K=\begin{pmatrix} f_u & 0 & u_0 \\ 0 & f_v & v_0 \\ 0 & 0 & 1 \end{pmatrix}$,which were the inner parameter of the camera. R was the rotation matrix, t is the translation vector.

As R and t was on the Lie Groups, in order to simplify the calculating process, map them to lie algebra, the camera's position in the lie algebra was named ξ ,so

$$\mathrm{T}=\begin{pmatrix} \mathrm{R} & \mathrm{t} \\ 0 & 1 \end{pmatrix}=\exp(\hat{\xi}) \tag{6}$$

the formula(5)might transferred into：

$$\mathrm{p}_i\,'=\begin{pmatrix} u_i\,' \\ v_i\,' \\ 1 \end{pmatrix}==\frac{1}{Z_i\,'}K\exp(\hat{\xi})\,\mathrm{P}_i \tag{7}$$

The gray value of the p_i pixel point was $\mathrm{I}_i(\mathrm{p}_i)$ and the gray value of the p'_i pixel point was $\mathrm{I}_i\,'(\mathrm{p}_i')$ the distance between two gray pixel was

$$e_i = \mathrm{I}_i(\mathrm{p}_i) - \mathrm{I}_i\,'(\mathrm{p}_i') \tag{8}$$

$$\begin{aligned}\mathrm{T}(\xi)=&\min_{\xi}\sum_{i=1}^{N}\| \mathrm{I}_i(\mathrm{p}_i) - \mathrm{I}_i\,'(\mathrm{p}_i')\|_2^2\\ &=\sum_{i=1}^{N}\| \mathrm{I}_i(\mathrm{AP}_i) - \mathrm{I}_i\,'(\mathrm{A}\exp(\hat{\xi})\,\mathrm{P}_i)\|_2^2\end{aligned} \tag{9}$$

The A as a parameter was $\mathrm{A}=\frac{1}{\mathrm{Z}_i}K$.

In the Lie Group, we assume the moving and rotating of the camera as $\Delta\xi$,so equation as fallow:

$$\begin{aligned}e_i(\xi+\Delta\xi) &= \mathrm{I}_i(\frac{1}{Z_i}\mathrm{KP}_i) - \mathrm{I}_i'(\frac{1}{Z_i\,'}\mathrm{K}\exp(\Delta\hat{\xi})\exp(\hat{\xi})\,\mathrm{P}_i)\\ &= \mathrm{I}_i(\frac{1}{Z_i}\mathrm{KP}_i) - \mathrm{I}_i'(\frac{1}{Z_i\,'}\mathrm{K}\exp(\hat{\xi})\,\mathrm{P}_i + \frac{1}{Z_i\,'}\mathrm{K}\,\Delta\hat{\xi}\exp(\hat{\xi})\,\mathrm{P}_i)\end{aligned} \tag{10}$$

where $\mathbf{w} = \Delta\hat{\xi}\exp(\hat{\xi})\,\mathrm{P}_i$, $\mathbf{u} = \frac{1}{Z_i\,'}Kw$,and we could get the first order Taylor expansion of the equation(10). The result as follow:

$$\begin{aligned}e_i(\xi+\Delta\xi) &= \mathrm{I}_i(\frac{1}{Z_i}\mathrm{KP}_i) - \mathrm{I}_i'(\frac{1}{Z_i\,'}\mathrm{K}\exp(\hat{\xi})\,\mathrm{P}_i + \mathbf{u})\\ &\approx \mathrm{I}_i(\frac{1}{Z_i}\mathrm{KP}_i) - \mathrm{I}_i'(\frac{1}{Z_i\,'}\mathrm{K}\exp(\hat{\xi})\,\mathrm{P}) - \frac{\partial \mathrm{I}_i\,'}{\partial u}\frac{\partial u}{\partial w}\frac{\partial w}{\partial \Delta\xi}\Delta\xi\\ &= e_i(\xi)\text{-}\frac{\partial \mathrm{I}_i\,'}{\partial u}\frac{\partial u}{\partial w}\frac{\partial w}{\partial \Delta\xi}\Delta\xi\end{aligned} \tag{11}$$

where $\frac{\partial \mathrm{I}_i'}{\partial u}$ was the gradient information of pixel. $\frac{\partial \mathbf{u}}{\partial \mathbf{w}}$ can be expressed by the equation(12),and $\frac{\partial \mathbf{w}}{\partial \Delta\xi}$ can be expressed by the equation(13).

$$\frac{\partial \mathbf{u}}{\partial \mathbf{w}} = \begin{pmatrix} \frac{\partial u}{\partial \mathrm{X}} & \frac{\partial u}{\partial \mathrm{Y}} & \frac{\partial u}{\partial \mathrm{Z}} \\ \frac{\partial v}{\partial \mathrm{X}} & \frac{\partial v}{\partial \mathrm{Y}} & \frac{\partial v}{\partial \mathrm{Z}} \end{pmatrix} = \begin{pmatrix} \frac{f_x}{Z} & 0 & -\frac{f_x\mathrm{X}}{Z^2} \\ 0 & \frac{f_y}{Z} & \frac{f_y\mathrm{Y}}{Z^2} \end{pmatrix} \tag{12}$$

$$\frac{\partial \mathbf{w}}{\partial \Delta\xi} = \begin{pmatrix} \mathrm{I} & \text{-}\hat{\mathbf{w}} \end{pmatrix} \tag{13}$$

As the $\frac{\partial \mathbf{u}}{\partial \mathbf{w}}$ and $\frac{\partial \mathbf{w}}{\partial \Delta\xi}$ were related to the point of space only but had nothing to do with the image,so we could calculate the equation.

$$\frac{\partial \mathbf{u}}{\partial \Delta\xi} = \begin{pmatrix} \frac{f_x}{Z} & 0 & -\frac{f_x\mathrm{X}}{Z^2} & -\frac{f_x\mathrm{XY}}{Z^2} & f_x + \frac{f_x\mathrm{X}^2}{Z^2} & -\frac{f_x\mathrm{Y}}{Z^2} \\ 0 & \frac{f_y}{Z} & -\frac{f_y\mathrm{Y}}{Z^2} & -f_y - \frac{f_y\mathrm{Y}^2}{Z^2} & \frac{f_y\mathrm{XY}}{Z^2} & \frac{f_y\mathrm{Y}}{Z} \end{pmatrix} \tag{14}$$

In summary,we could get the Jacobian matrix in Lie algebra of the relative deviation.

$$\mathbf{J}=-\frac{\partial \mathrm{I}_i{}'}{\partial \mathbf{u}}\frac{\partial \mathbf{u}}{\partial \Delta \xi} \tag{15}$$

Then Gauss Newton method was selected to iterative calculate the increment, so the position estimation was updated.

4.Experiments and Analysis

In this thesis,we studied the visual odometry location technology based on RGB-D sensor deeply and analysed the major parts of the algorithm.Then,starting with Fast Odometry from Vision based on features and Dense Visual Odometry based on direct method,they were verified and analysed in different kinds of environment.

4.1 Experiment platform

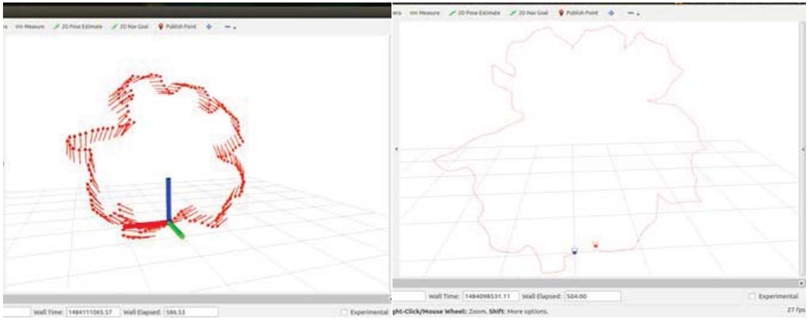
The test platform is equipped with I3 processor,2.2GHz frequency and 4GB RAM.We develop our localization system using ROS Indigo,OpenCV and C++ language.As a popular RGB-D sensor,the kinect V1 is selected to capture experimental image data.

4.2 Visual odometry estimation experiment based on RGB-D

(1)The experiment in office environment

In order to analyse the accuracy of two algorithms to estimate position， the authoritative freiburg2_desk data sets in TUM was chosen and the image data was recorded by KinectV1 at full velocity (30HZ), the resolution of the sensor is 640x480. The data set contained a color image, depth image. It was recorded in a typical office scene.

RVIZ was a powerful visualization tool in ROS, the motion estimation based on the FOVIS algorithm and DVO algorithm could showed by the RVIZ, the result as shown in Figure 2 :



FOVIS motion estimation DVO motion estimation

Fig 2 Motion estimation in office environment

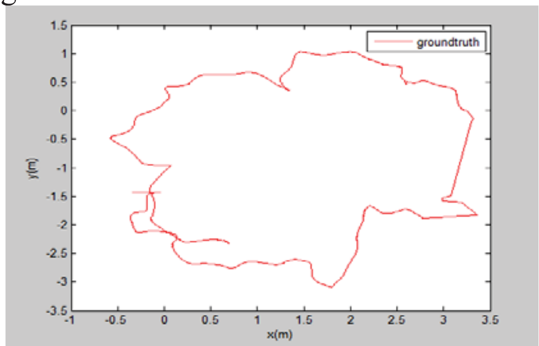
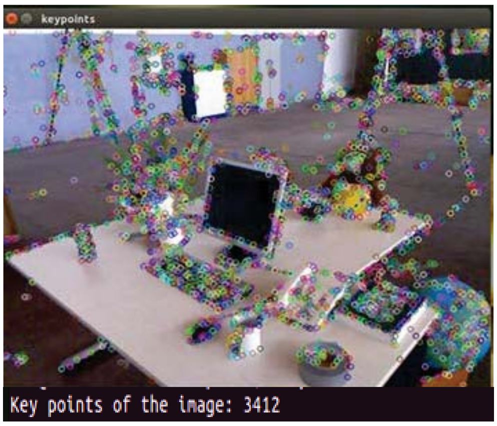


Fig.3 Groundtruth trajectory of freiburg2_desk environment

The TUM data set provided the movement information which was captured by eight high speed tracking camera(100Hz), and the true trajectory was drawn in MATLAB, the result was showed in figure 3.



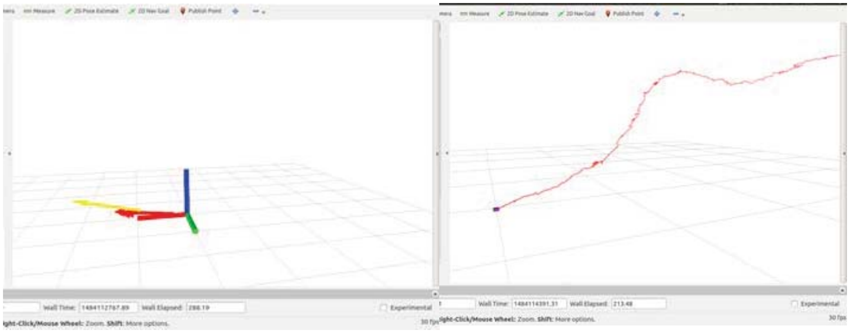
Key points of the image: 3412

Fig.4 Feature extraction

The light and feature of the office scene was rich， a frame was selected to extract the feature by the FAST algorithm and result showed in figure4. The feature mount extracted in the frame was 3412. In the visual feature riched environment, the FOVIS algorithm based on feature was more accurate than DVO direct method.

（2）The experiment in ground environment

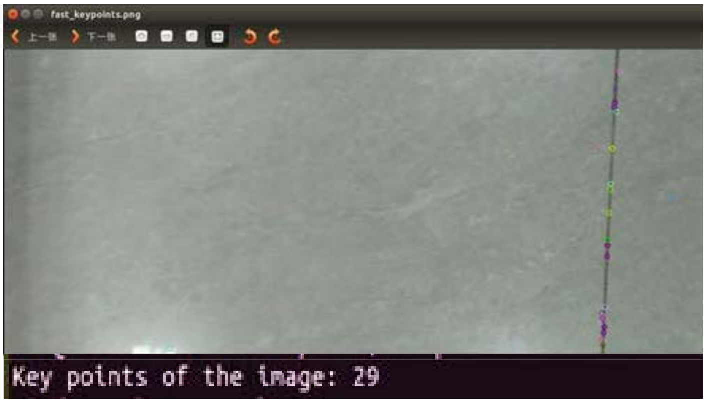
What we have considered in the actual scenario is that the failure or error estimation of odometry will be more capable to influenced the performance of location than inexact conditions.Compared with accuracy ,robustness is relatively more important.In contrast to the office environment with rich features,we captured images in smooth ground environment.There are few reference objects can be used for motion estimation in this simple environment.The camera's trajectories are shown in Fig.5 which were estimated by FOVIS algorithm and DVO algorithm respectively.



FOVIS motion estimation

DVO motion estimation

Fig.5 Motion estimation in floor environment



Key points of the image: 29

Fig.6 Feature extraction

It can be seen from Fig.5 that the motion estimation result of DVO is better than FOVIS algorithm in ground environment with good light. As compared to the complex office environment, the ground environment is too simple.We choosed an image from this experimental datasetas to extract features.We obtained 29 effective features from this image.The results are shown in Figure 6.FOVIS algorithm based on features only worked for a short distance.It was invalid without enough information of visual feature.So in this degenerate environment,the algorithm based on features is of bad robustness.Instead,the direct method estimated motion very well because of using the whole pixel information.

4.3 Analysis and summary

The experiment showed that the FOVIS based on RGB-D could achieve accurate motion estimation in light texture feature riched environment. But there were disadvantage of FOVIS algorithm, a lot of useful information in the image could be discarded during the process of image feature extraction. So the result might not accurate. The DVO algorithm of direct method without feature extraction， was based on pixel information to estimate the motion directly. The motion was estimated by solving the pixel gradient, the robustness of the algorithm was excellent even in a poor condition, such as the degradation of optical feature information.

5. Conclusions

In this paper, the RGB-D visual odometry localization technology was studied and the main aspects of the algorithm was analyzed. Then the FOVIS algorithm based on RGB-D and DVO algorithm,a direct method, was taken as an example, and the experiment was carried out in various of environment. The performance of the direct method was more accurate in the environment of the optical feature degradating. And the visual odometry estimation algorithm based on feature method was more accurate in the scene with rich visual features. After a series of analyses we could conclude clearly that different algorithms could achieve accurate location in the specific environment. However, there is no algorithm to realize fast and accurate location in any environment，they have their own advantages and disadvantages.So, we could comprehensive various location technology about the robot location. Selecting different visual range estimation algorithm based on the mount of environmental feature realize the localization of real-time, high accuracy and strong robustness.

References

[1] Chenavier F., Crowley J.L,Position estimation for a mobile robot using vision and odometry.IEEE International Conference on Robotics and Automation, Proceedings. IEEE,Piscataway, NJ, USA:IEEE,pp.12-14,1992.

[2] Betke M, Gurvits L.Mobile robot localization using landmarks.IEEE Transactionson Robotics and Automation, , 13(2),pp.251-263, 1997.

[3] Thrun S.Simultaneous localization and mapping.Springer Tracts in Advanced Robotics, 38(3),pp.13-41, 2008.

[4] Smith R, Self M, Cheeseman P,Estimating uncertain spatial relationships in robotics,Autonomous Robot Vehicles,Springer:New York,pp.435-461, 1990.

[5] Konolige K, Agrawal M, Solà J. Large-Scale Visual Odometry for Rough Terrain.Robotics Research. Springer:Berlin Heidelberg,pp.201-212,2010.

[6] Bonin-Font F,Ortiz A,Oliver G.Visual navigation for mobile robots:A survey.Joumal of Inteuigent&Robotic Systems, 53(3),pp.263-296,2008.

[7] Kerl C,Sturm J,Cremers D.Robust odometry estimation for RGB-D cameras.IEEE International Conference on Robotics and Automation,IEEE, Karlsruhe, Germany , pp.3748-3754, 2013.

[8] Pomerleau F, Colas F, Siegwart R,Comparing ICP variants on real-world data sets. Autonomous Robots, 34(3),pp.133-148,2013.

[9] Nister D,Naroditsky O,Bergen J,Visual odometry.Computer Vision and Pattern Recognition,Proceedings of the 2004 IEEE Computer Society Conference on.DBLP, Washington, D.C. USA,pp.652-659, 2004.

[10] Whelan T, Johannsson H, Kaess M,Robust real-time visual odometry for dense RGB-D mapping.IEEE International Conference on Robotics & Automation. IEEE, Karlsruhe, Germany ,pp.5724-5731,2013.

[11] Dryanovski I, Valenti R G, Xiao J. Fast visual odometry and mapping from RGB-D data,IEEE International Conference on Robotics and Automation. IEEE, Karlsruhe, Germany 2305-2310,2013.

[12] Fischler M A, Bolles R C,Random Sample Consensus: A Paradigm for Model Fitting with Applications To Image Analysis and Automated Cartography, Communications of the Acm, 24(6),pp.381-395, 1981.

[13] Irani M,Anandan P.About Direct Methods,Lecture Notes in Computer Science, Springer-Verlag:Berlin,pp.267-277,1999.

References

Reference to journal papers :期刊论文

[1] Heider, E.R.& D.C.Oliver. The structure of color space in naming and memory of two languages [J]. Foreign Language Teaching and Research, 1999, 16(3): 62-67.

Reference to a book: 书

[2] Gill, R. Mastering English Literature [M]. London: Macmillan, 1985: 42-45.

Reference to conference papers: 会议论文

[3] [10] Almarza, G.G. Student foreign language teacher's knowledge growth [A]. In D.Freeman and J.C.Richards (eds.). Teacher Learning in Language Teaching [C]. New York: Cambridge University Press. 1996. pp.50-78.

Reference to conference papers: 网页内容

[4] Information on http://www.weld.labs.gov.cn