IMAGE RECOGNITION FOR MOBILE APPLICATIONS

Jimmy Addison Lee and Kin Choong Yow

Nanyang Technological University, Singapore
School of Computer Engineering
Block N4, Nanyang Avenue, Singapore 639798

ABSTRACT

Our paper presents a system for efficient recognition of landmarks taken from camera phones. Information such as tutorial rooms within the captured landmarks is returned to user within seconds. The system uses a database of multiple viewpoint's images for matching. Various navigational aids and sensors are used to optimize accuracy and retrieval time by providing complementary information about relative position and viewpoint of each query image. This makes our system less sensitive to orientation, scale and perspective distortion. Multi-scale approach and a reliability score model are proposed in this application. Our system is validated by several experiments in the campus, with images taken from different resolution's camera phones, positions and times of day.

Index Terms— Image recognition, reliability score, camera phones, mobile information guide

1. INTRODUCTION

With the advances in mobile technologies and more than a billion of mobile subscribers globally reported by Advisor PC [1], mobile devices nowadays have more than just communication capabilities. Most mobile phone these days are equipped with built-in camera and multimedia messaging capability. Present day mobile networks can also support high speed data transfer. With these, a lot of new applications can be made possible. One example is mobile information guide. A user equipped with a camera phone can use this device to help him find information about his surroundings. The idea is that the user can transmit a photograph of his surroundings to a central computer system. With various navigational aids and sensors, images acquired by the camera phones can provide additional means for determining the position and viewpoint of users. This approximate knowledge can be used to retrieve stored images in the database in order to determine pose. Information can then be mapped to the query image and transmitted back to the user.

In an article by Markoff et al. [2], it introduces new technology phones in Japan that combine satellite-based navigation with an electronic compass to provide a new dimension of orientation. Our application makes use of this similar concept and further verifies the matching results in our reliability score model.

1.1. Related work

Several applications have made use of camera phone devices. Fockler et al. [3] and Fleck et al. [4] have presented tour-guide systems that targeted on museums. Camera phones and on-device object recognition are used. But the method is not well suited for outdoor applications and the recognition rate decreases with an increasing number of objects.

A more similar application to ours is proposed in Robertson et al. [5] and Cipolla et al. [6]'s papers. They designed a system to allow users to navigate in an urban environment using camera phones. The system facilitates efficient determination of the pose of a query view by reference to a database of views of building facades. The limitation is that some buildings or parts of the buildings are similar and the system might not be able to distinguish between some viewpoints. Furthermore, conducting the two view matching between the query view and image database is slow.

1.2. Approaches and framework

In this paper we describe the techniques used in the application processes. The Application framework is presented in Figure 1. The query image provides the approximate position and orientation to the central computer system with the use of integrated Global positioning sensor (GPS) and compass on the mobile phone. This greatly minimize the database retrieval time by only selecting images in the image database that are within the region and viewpoint. Quality corner features are formed by computing the curvature in multi-scale, and used for images correlation.

Next, a reliability score model is introduced to calculate the reliability of each pair of correct corners match in the 2 images by combining a number of factors: the results of correlation, distance voting and spatial relations matching. The few most reliable pairs of corners are chosen for pose recovery. The following Sections will discuss these processes in more details.

Thanks to Nanyang Technological University, Singapore for funding.

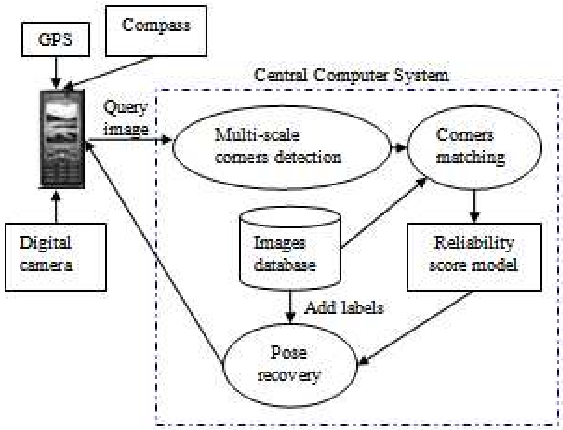


Fig. 1. Application framework

2. DETECTION AND MATCHING TECHNIQUES

After receiving the query image from the user, the first process is to extract corner features from this image, and then retrieve selected images from the image database for matching. Database retrieval, corners detection and images correlation processes will be discussed in this Section.

2.1. Database retrieval

Our image database consists of 8 angles for each position (North, North-East, East, South-East, etc.), and 10 meters interval for each position. A set of example is shown in Figure 2. 3 sets of images with the closest coordinate positions with the query image will be retrieved for images correlation. These images will be further reduced by eliminating those that are out of orientation from the query image. For example if the query image is taken at 10 degree which is closest to North (North is 0 degree), only those images that are North oriented will be retrieved from these 3 sets of images.

Consequently, the query image and the retrieved images from the image database will have very close orientation. This greatly compromises the orientation, scale and perspective distortion problems during images correlation process.

2.2. Multi-scale corners detection

We match the images by their sets of quality corners. The quality of each corner feature is measured by its repeatability. A corner feature that appears repeatedly in multiple scale is considered as a consistent corner feature. What we do is to re-scale the image several times to do the corner detection and pick the consistent corner features. An example of the multi-scale corners detection's result is shown in Figure 3.

We use the algorithm derived by He et al. [7] to do our corner detection. This robust corners detector first uses an adaptive local curvature threshold instead of a single global threshold as in the original and enhanced Curvature Scale Space methods. Secondly, the angles of the corner candidates are checked in a dynamic region of support for eliminating falsely detected corners.

Fig. 2. A set of images from image database: 8 angles in a position. North (starts from top left to right), North-East, East, South-East, South, South-West, West and North-West).

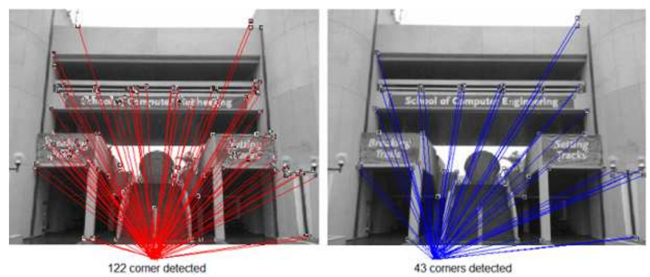


Fig. 3. Left: 122 corners in a single scale. Right: 43 strong corners remain after multi-scale (after 3 iterations).

2.3. Images correlation

As mentioned in Section 2.1, 3 sets of images with the closest coordinate positions and orientations with the query image will be retrieved from the image database for images correlation. Normalized zero mean cross-correlation ncc which is less sensitive to illumination change is used:

$$ncc = \frac{\sum(p-\bar{p})(q-\bar{q})}{\sqrt{\sum(p-\bar{p})^2\sum(q-\bar{q})^2}} \quad (1)$$

where p and q are image patches around the corners, each in different images. $\bar{p}$ and $\bar{q}$ are their respective means. $result_c(i)$ is the pair with the highest ncc result, with respect to the corner i in an image. There will certainly be a number of correct and incorrect pairs of match. A reliability score model will be introduced in the next Section to deal with the reliability of each pair of corners match.

3. RELIABILITY SCORE MODEL

We proposed a reliability score model to verify the reliability of each pair of corners match, by calculating the reliability score of each correct match. This is done by adding the correlation's result with the distance voting's result and spatial relations matching's result.

3.1. Distance voting

Distance voting is the Euclidean difference of each pair of corresponded corners in the 2 images with every other pairs of matched corners. Since the 2 images have similar positioning and orientation, the distance between these pairs with the other neighboring pairs should not be too much in difference. Each pair with a small Euclidean distance difference will win a vote. Each vote has its different weighage (e.g. a pair between 5-10 pixels in Euclidean distance difference will win 0.5 vote, while another pair below 5 pixels in Euclidean distance difference will win 1 vote). The n numbers of accumulated votes v_i of each pair i is defined in a set V as:

$$V = \{v_1, v_2, v_3, \ldots, v_n\} \tag{2}$$

and the distance voting's result of each pair i after normalization will be:

$$result_d(i) = \frac{v_i}{max(V)} \tag{3}$$

Figure 4 shows each vote by drawing a line between the pair.

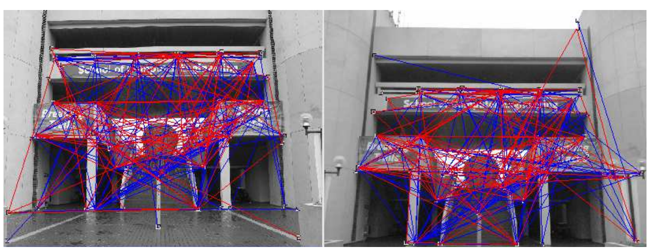


Fig. 4. Distance voting. The left and right images are the image captured with camera phone and image database respectively. Red line: Distance difference between 5-10 pixels. Blue line: Distance difference $<$ 5 pixels.

3.2. Spatial relations matching

Spatial relations matching assesses the reliability of each pair of matched corners in the 2 images based on the spatial relations with all other pairs. If all the other pairs have angles that are very similar, then the spatial relations matching's result for this pair is high. Each angle of the pair is formed by the line made with the pair and the x-axis:

$$\theta_i = \tan^{-1}\frac{y_{i2} - y_{i1}}{x_{i2} - x_{i1}} \tag{4}$$

where $\theta_i \in [0°, 360°)$, i is the index of a matched pair, $n-1$ is the total number of matched pairs, and $0 \le i \le n-1$. The number (1 or 2) after i specify which image it is, the query image (1) or the database image (2). Example y_{i1} is the y-axis of ith point in the query image.

We have calculated the angle θ_i for each matched pair. We will now do voting on each pair. If the angle θ_i of ith pair is very similar ($< 10°$) to another pair, it wins a vote. Each vote is accumulated to form a total number of vote for that particular pair. The equation is as follows:

$$count_k = \sum_{j=1, k\neq j}^{n} f(\theta_k - \theta_j) \tag{5}$$

$$f(\theta_k - \theta_j) = \begin{cases} 1 & \text{if } |\theta_k - \theta_j| \le 10° \text{ or } |\theta_k - \theta_j| \ge 350° \\ 0 & \text{otherwise} \end{cases}$$

We will then normalize it as follows:

$$result_s(k) = \frac{count_k}{max(\{count_1 \ldots count_{n-1}\})} \tag{6}$$

Finally, we combine this result with the previous distance voting and correlation's results:

$$score(i) = \frac{result_c(i) + result_d(i) + result_s(i)}{3} \tag{7}$$

The reliability score table will look like the example as shown in Table 1 below:

Im_1's corners i	Im_2's corners i	$score\ i$
(375, 346)	(446, 337)	0.9543
(103, 387)	(178, 382)	0.9278
(354, 252)	(419, 245)	0.9263
(443, 433)	(515, 420)	0.9018
(324, 402)	(391, 393)	0.8901
(445, 209)	(514, 199)	0.8823
(284, 501)	(357, 489)	0.8396
(352, 445)	(420, 436)	0.7626
(307, 454)	(375, 445)	0.7590
(122, 554)	(199, 543)	0.6953

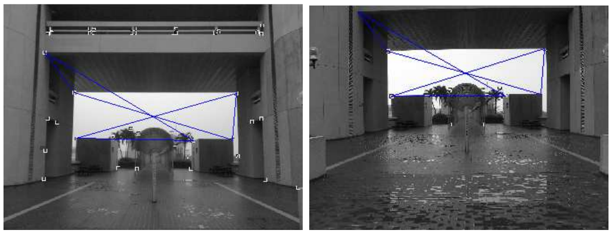
Table 1. Top 10 scores. 1st column and 2nd column: coordinates of corner points in image 1 (Im_1) and image 2 (Im_2) respectively. 3rd column: Reliability of each matched pair i.

3.3. Selecting corner pairs

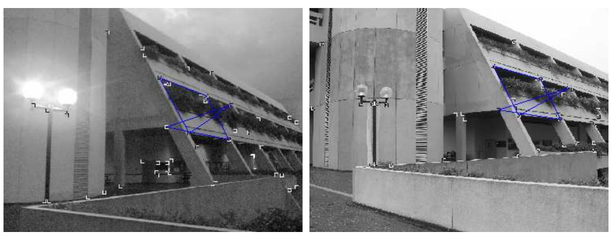
We have to select 4-8 pairs from the reliability score table to do pose recovery. In order to decide how many pairs to select, we have to refer to the score in the score table. After several test cases, we defined a benchmark score of 0.8 in order to rate the corner pair as reliable. For example in Table 1, we will pick the top 7 pairs, as the 8th pair has a score of 0.7626, which is below 0.8. Figure 5 shows some selected corners in different pairs of images.

As we can see from Figure 5, the corners selected are corresponded in each pair of images. The system can notably deal with illumination change as shown in Figure 5c. The final step is to use these selected pairs to do pose recovery and information mapping. We will see this in the next Section.

(a) 8 reliable corners selected



(b) 6 reliable corners selected



(c) 6 reliable corners selected

Fig. 5. Corresponded corners selected (lines drawn) - Left side: Images captured with camera phone. Right side: Images from image database.

4. POSE RECOVERY

At least 4 corresponded points are needed to do pose recovery between the views from planer homographies. This technique has been used by Robertson et al. [5] in their paper. The homographies equation is as followed:

$$\begin{bmatrix} x_i' \\ y_i' \\ 1 \end{bmatrix} \cong \begin{bmatrix} h_{00} & h_{01} & h_{02} \\ h_{10} & h_{11} & h_{12} \\ h_{20} & h_{21} & h_{22} \end{bmatrix} \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} \tag{8}$$

Information can be mapped to the query image from the image database once this is done, as shown in Figure 6.

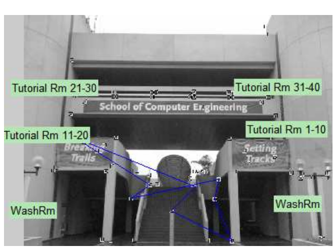


Fig. 6. Information mapped to query image after homographies using 8 corner points.

5. EVALUATION

We have constructed a database of views by photographing one of the spines in our campus, classified by their coordinates and orientations (8 orientations at 10 meters interval). Our image database comprises of over 300 images. On the other hand, we have taken 54 query images at various positions, times of the day, and resolutions. Each query took no longer than 10 seconds on a 1.83GHz notebook computer. Overall, 48 out of 54 queries were matched correctly.

6. CONCLUSION

We have described a prototype that allows users to find information about his surroundings. Multi-scale approach is used for extracting strong corners from the images. After performing correlation to find the corresponded pairs of corners in the 2 images, a reliability score model is proposed to calculate the reliability of each matched pair. We have made use of the GPS and compass to optimize accuracy and retrieval time by providing complementary information about relative position and viewpoint of each query image taken. This makes our application less sensitive to orientation, scale and perspective distortion. Several experiments have been performed and the results show that this system works well.

7. REFERENCES

[1] Advisor PC, “Mobile users to reach 1 billion,” vol. Issue 91, Spring, http://www.pcadvisor.co.uk.

[2] J. Markoff and M. Fackler, “With a cellphone as my guide,” June 2006.

[3] P. Focklers, T. Zeidler, B. Brombach, E. Bruns, and O. Bimber, “Phoneguide: Museum guidance supported by on-device object recognition on mobile phones,” in *MUM’05*, 2005.

[4] M. Fleck, M. Frid, T. Kindberg, E. O’Brien-Strain, R. Rajani, and M. Spasojevic, “From informing to remembering: Ubiquitous systems in interactive museums,” in *IEEE Pervasive Computing*, 2002, vol. 1.

[5] D. Robertson and R. Cipolla, “An image-based system for urban navigation,” in *BMVC*, 2004.

[6] R. Cipolla and D. Robertson, “3d models of architectural scenes from uncalibrated images and vanishing points,” in *ICMCS*, 1999, vol. 1, pp. 25–31.

[7] X. C. He and N. H. C. Yung, “Curvature scale space corner detection with adaptive threshold and dynamic region of support,” in *ICPR’04*, 2004, vol. 2, pp. 791–794.