WHERE WAS THE PICTURE TAKEN: IMAGE LOCALIZATION IN ROUTE PANORAMAS USING EPIPOLAR GEOMETRY

Saad M. Khan, Fahd Rafi, Mubarak Shah

Department of Computer Science, University of Central Florida, USA

ABSTRACT

Finding the location where a picture was taken is an important problem for a variety of applications including surveying, interactive traveling and homeland security among others. The task becomes intractable though when the area under investigation reaches city/town size. The amount of data (pictures/videos) required to visually map a city, comprehensively, can be exhaustive for most search algorithms. In this paper we propose a novel method to effectively tackle this problem. The area is visually mapped as route panoramas that provide a compact yet comprehensive representation of the buildings and landmarks in the area. Given a query image taken at an arbitrary location in the area, we show that we can accurately recover the location of the camera by finding it's epipole in the route panorama of the scene. To this end we show that there exists a fundamental matrix between a route panorama and a perspective image of the same scene. The fundamental matrix is calculated using feature matches as correspondences between the query image and the route panorama.

1. INTRODUCTION

In this paper we address the problem of automatically locating the spot, in a large area like a city, where a picture was taken. For example if a person takes a picture with his/her cell phone in an outdoor setting, using only the image information, we would like to automatically determine the location where the picture was taken. The central hypothesis is that every location in a large area like a town or a city is distinguished by its peculiar landmarks and features. This is usually how people ascertain where a picture was taken by recalling the landmarks and features of the buildings present in the picture. Nevertheless it is a tough task for even humans and usually takes many years of exploring and repetitive viewing to confidently 'know' a city.

This problem is of particular interest in the fields of surveying large areas, interactive traveling and homeland security operations like analyzing spy photos. The first task is to build a comprehensive database of images taken at locations distributed over the entire area. In the case of areas the size of cities, the amount of data necessary to visually map is prohibitively large. There is the added problem of redundancy in the collected data due to overlapping views of the same scene. These problems compound the difficulties faced in visual recall and can render the task intractable.

An efficient way of visually mapping large areas is to use route panoramas [1]. A route panorama is essentially a mosaic of the scene, created as the camera moves along a route. While making route panoramas we keep a track of which panorama corresponds to which section of the city. Given a picture taken at any arbitrary spot, visual recall is done by performing feature matching between the given image and the database of route panoramas. SIFT features [2] can be used for this purpose as these are invariant across a substantial range of affine distortion.

To verify the correctness of the matches, and to reliably localize the position where the query image was taken we introduce a geometrical constraint. We show that there exists a fundamental matrix between a route panorama and a normal perspective image of the same scene much like the fundamental matrix constraint between two stereo perspective images. Using feature matches as correspondences we calculate the fundamental matrix that gives us the epipole of the query image in the best matching route panorama. The epipole is the projection in the route panorama of the query image's camera center and provides a reliable estimate of the location where the query image was taken. On large scales like a city, this estimate is reasonably precise.

The remainder of this paper is organized as follows. In section 2 we describe route panoramas in detail and justify their selection as a medium ideally suited for our problem. Section 3 discusses our feature matching strategy. In section 4 we derive the fundamental matrix constraint and how the epipole can be determined. In section 5 we present our algorithm and in section 6 we analyze the experimental results. The paper is concluded in section 7.

2. ROUTE PANORAMA

A route panorama is synonymous to an image belt. It is a single panoramic view of the entire scene along a route. To

This material is based upon the work funded in part by the U.S. Government. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the U.S. Government.

Fig. 1. An example route panorama. The ortho-perspective projection means closer objects like cars, trees and poles get squeezed in, while important landmarks, like buildings that are further away are adequately captured.

create a route panorama the first step is to scan scenes continuously with a camera. For each frame only a vertical pixel line at a fixed position is considered, the rest is discarded. Finally by pasting these consecutive slit views together to form a long, seamless 2D image belt a route panorama is created [1]. This viewing scheme is an ortho-perspective projection of scenes: orthogonal toward the camera path and perspective along the vertical direction. Through the coarse of this paper the terms route panorama and ortho-perspective image will be used interchangeably.

Compared with other approaches to model a route using graphics models [3] route panoramas have an advantage in capturing scenes. They don't require taking discrete images by manual operation or texture mapping onto geometric models. A route panorama can be ready after driving in town for a while with a camera mounted on the vehicle. It yields a continuous image scroll that with other image stitching or mosaic approaches, [4] in principle, is impossible to realize due to changes in depth. A route panorama requires only a small fraction of data compared to a video sequence. If we pile a sequence of video frames along the time axis, into a spatio-temporal volume. The route panorama comprises pixel lines in consecutive image frames, which correspond to a 2D data sheet in the spatiotemporal volume. Ideally, if the image frame has a width w (in pixels), a route panorama only has $1/w$ of the data size of the entire video sequence. Route panoramas neglect redundant scenes in the consecutive video frames. The missing scenes are objects under occlusion when exposed to the slit. For an in-depth analysis of these and other properties of route panoramas the interested reader is directed to the paper by Zheng et al [1].

The capability of route panoramas to comprehensively represent large outdoor environments in a compact and continuous manner, makes them an ideal medium to visually map and index large areas like a town or a city. While creating route panoramas we keep a track of which panorama corresponds to which section of the city. This could be achieved by annotating route panoramas with the street or road names that they capture or by using GPS data if available. Figure 1 shows an example route panorama from our experiments. Once a large area has been mapped with route panoramas the next task is to visually index the data by its distinguishing features and landmarks. This process is similar to the human cognitive process wherein a familiar building or location is recalled by its peculiar features and markings.

3. FEATURE MATCHING

Scale Invariant Feature Transform (SIFT) features [2] can be used to perform reliable matching between different views of an object or scene. The features are invariant to image scale and rotation, and provide robust matching across a a substantial range of affine distortion, change in 3D viewpoint, addition of noise, and change in illumination. The features are highly distinctive, in the sense that a single feature can be correctly matched with high probability against a large database of features from many images.

For image matching and recognition, SIFT features are extracted from the query image and the route panorama. Candidate matches are found based on the Euclidean distance between their feature vectors. The descriptors used in SIFT are highly distinctive, which allows a single feature to find its correct match with good probability in a large database of features. The matches obtained from this step are used as correspondences, to calculate the fundamental matrix that encapsulates the geometrical relationship between a route panorama and a normal perspective image of the same scene.

4. FUNDAMENTAL MATRIX FOR LOCALIZATION

The problem of determining the relative camera placement of two or more pinhole cameras and consequent determination of pinhole cameras has been extensively considered [5, 7]. If $\{(u_i, u_i')\}$ is a set of match points in a stereo pair, F, the fundamental matrix is defined by the relation $u_i'^T F u_i = 0$ for all i. A fundamental matrix was shown to exist between two ortho-perspective cameras (route panoramas)[6]. In this section we show that a similar relation exists between a ortho-perspective image and a perspective image, i.e between a route panorama and the image of the same scene taken with a normal camera. Using this formulation we show that the epipole of the query (perspective) image can be determined, which gives a good estimate of where the picture was taken with respect to the route panorama.

Consider a point $X = (x, y, z)^T$ in space as viewed by an ortho-perspective camera and perspective camera with camera matrices M and M' respectively. Let the images of the two points be $u = (u, v)^T$ and $u' = (u', v')^T$. This gives a pair of equations:

$$(u, wv, w) = M(x, y, z, 1)^T, \tag{1}$$

Fig. 2. The picture on the left is the query image, the one on the right is the matched route panorama. Features detected and matched between the two images are marked with circles that are color coded (e.g the blue circle in the query image matches the blue circle in the route panorama). The fundamental matrix is calculated with these matches. Mapping feature points from the query image to the route panorama using the fundamental matrix creates epipolar hyperbolas that intersect at the epipole. It can be seen that the epipole accurately determines where in the route panorama the query image was taken.

$$(w'u', w'v', w') = M'(x, y, z, 1)^T. \tag{2}$$

Note that the form of equation 1 means the camera projection is perspective in the vertical direction but orthographic in the horizontal direction. This pair of equations can be written in the following form:

$$\begin{bmatrix} m_{11} & m_{12} & m_{13} & m_{14}-u & 0 & 0 \\ m_{21} & m_{22} & m_{23} & m_{24} & v & 0 \\ m_{31} & m_{32} & m_{33} & m_{34} & 1 & 0 \\ m'_{11} & m'_{12} & m'_{13} & m'_{14} & 0 & u' \\ m'_{21} & m'_{22} & m'_{23} & m'_{24} & 0 & v' \\ m'_{31} & m'_{32} & m'_{33} & m'_{34} & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \\ w \\ w' \end{bmatrix} = 0. \tag{3}$$

The 6x6 matrix in (3) will be denoted $A(M, M')$. Considered as a set of linear equations in the variables x, y, z, w, and w' and constant one, this is a set of six homogeneous equations in six unknowns (imagining one to be an unknown). If this system is to have a solution, then det $A(M, M') = 0$. This condition gives rise to an equation $p(u, v, u', v') = 0$, where the coefficients of p are determined by the entries of M and M'. The polynomial p is called the fundamental polynomial corresponding to the two cameras. Because of the particular form of (3) there are no terms in $u^2, u'^2, v^2,$, v'^2 and $u'v'$in the fundamental polynomial. Consequently, there exists a 4x3 matrix F such that $p(u, v, u', v') = 0$ may be written:

$$(u, uv, v, 1)F(u', v', 1)^T = 0. \tag{4}$$

The matrix F will be called the fundamental matrix corresponding to the ortho-perspective camera and the perspective camera pair M, M'. Matrix F is just a convenient way to display the coefficients of the fundamental polynomial. Since the entries of F depend only on the two camera matrices, M and M', (4) must be satisfied by any pair of corresponding image points (u, v) and (u', v'). The same basic proof method used above was used to prove the existence of the fundamental matrix between two ortho-perspective cameras [6] and for pinhole cameras [7]. It is seen that if either M or M' is replaced by an equivalent matrix by multiplying the last two rows by a constant c, then the effect is to multiply det$A(M, M')$, and hence the fundamental polynomial p and matrix F by the same constant c. Consequently, the fundamental matrix is defined only upto a constant scale factor and contains no more than 11 degrees of freedom. Given a set of 11 or more image-to-image correspondences obtained from feature matching, a linear least squares solution to the matrix F can be determined.

Outliers can now be removed by checking for agreement between each image feature and the model. If fewer than 11 points remain after discarding outliers, then the match is rejected. As outliers are discarded, the least-squares solution is re-solved with the remaining points, and the process iterated till convergence. The fundamental matrix obtained determine's the epipole of the query image, which is the projection of the query image's camera center in the route panorama.

If we plug in values of u', v' (from a match in the perspective image) and the fundamental matrix F in equation 4 we obtain the following form:

$$(u, uv, v, 1).(a, b, c, d)^T = 0. \tag{5}$$

This is an equation for a bilinear function with coefficients a, b, c and d that is satisfied by coordinates u, v. The plot of this function is an hyperbola in the route panorama that passes through the pixel (u, v). We call this the *epipolar hyperbola.* The geometric interpretation of this process is to project a ray through the pixel (u', v') in the perspective image, the image of this ray in the route panorama corresponds to the epipolar hyperbola. A line moving in depth is projected as a hyperbola in an ortho-perspective projection [1]. Taking several of these hyperbolas corresponding to different matches in the query image, we can find their point of intersection. This point is where the rays are emanating i.e the epipole of the query image. Figure 2 shows an example of the epipole formed by the intersection of epipolar hyperbolas.

Assuming that the query images are not taken at locations far off in depth from the path of the route panorama, the epipole is a good approximate of the query image's camera location relative to the route panorama. Since we already know which section of the city is mapped by the selected route panorama, we can now tell where on a particular road or street the query image was taken. On large scales like a city this is a reason-

ably precise solution.

5. ALGORITHM

As input to our algorithm we supply the database of route panoramas and the query image.

Require: Route panorama images for city stored
  for all Panorama images **do**
    Calculate feature matches of query image with panorama image
  end for
  Select best matching panorama image and set of matches M with the query image
  while Convergence in M is not reached **do**
    if $|M| < 11$ **then**
      Discard matches and Exit
    end if
    Calculate least square solution for Fundamental Matrix F using matches as correspondences u, v, u' and v' in equation 4.
    for all matches $i \in M$ **do**
      $Residue = mag([u_i\ v_i\ u_i v_i\ 1] \cdot F \cdot [u'_i\ v'_i\ 1]^T)$
      if $Residue > Threshold$ **then**
        Prune out match i And update M
      end if
    end for
  end while
  Calculate Epipole pixel location using F And **Return**

6. EXPERIMENTAL ANALYSIS

In order to evaluate the effectiveness of our method we compared the accuracy of our results with ground truth data. For each query image used in the evaluation we calculate its epipole in the best matching panorama using our algorithm. The ground truth projection of the query image's camera center in the route panorama is recorded while taking the picture and making the route panorama. Recall that a route panorama is an orthographic projection in the horizontal direction. If we annotate route panoramas with distance information by recording the distance from the start to the end of the route. Then the horizontal distance (in pixels) between any two points on a route panorama can be transformed into the world distance.

The error between the epipole and the ground truth projection of the camera center is transformed into world distance. Figure 3 shows a plot of the error in epipole for query images taken at various distances from the route of the route panorama. For distances between 2 and 80 meters the error in epipole varies between 2 and 14 meters. There are several sources of this error, including inaccuracies in feature matching and camera perturbation in route panorama. However, errors are within reasonable limits and on large scales like a city, the results are accurate and precise enough.

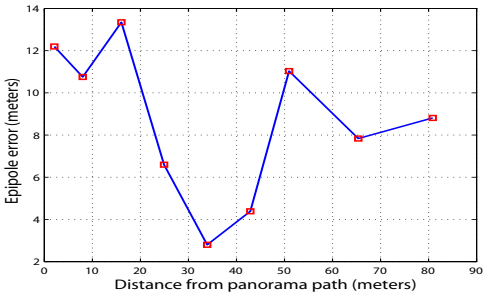

Fig. 3. A plot of error in the calculated query image location at various distances from the path of the route panorama.

7. CONCLUSIONS

In this paper we have proposed a method of automatically finding the location in a city where a picture is taken. We have exploited the potential of route panoramas to create a comprehensive visual repositories of large areas. To find the location of the spot where a query image is taken, we showed that there exists a fundamental matrix that uniquely determines it's epipole in the route panorama of the same scene. The epipole of the query image is a good estimate of it's camera center location with respect to the route panorama. If the route panoramas are annotated with data that links them to the appropriate sections of the city, using the epipole we can reliably state where the query image was taken. The applicability of our method makes it useful for a wide range of applications from surveillance to image based search.

8. REFERENCES

[1] Zheng J. Y., Digital Route Panorama, IEEE Multimedia, Vol.10, No.3. pp.57-68, 2003.

[2] D. G. Lowe, Distinctive Image Features from Scale-Invariant Keypoints, International Journal of Computer Vision, 60, 2 (2004), pp. 91-110.

[3] T. Ishida, Digital City Kyoto: Social Information Infrastructure for Everyday Life, Comm. ACM, vol. 45, no. 7, July 2002 .

[4] Z. Zhu, E.M. Riseman, and A.R. Hanson, Parallel- Perspective Stereo Mosaics, Proc. IEEE ICCV 2001.

[5] H.C. Longuet-Higgins, A Computer Algorithm for Reconstructing a Scene From Two Projections, Nature, vol. 293, pp. 133135, Sept. 1981.

[6] R. Gupta, R.I. Hartley, Linear Pushbroom Cameras, IEEE TPAMI, Vol. 19, No. 9, September 1997

[7] O. Faugeras and B. Mourrain, On the Geometry and Algebra of the Point and Line Correspondences Between N Images, IEEE ICCV 1995.