Structure Based Matching between Aerial and Map Images using Brightness- and Rotation-Invariant Curve Features

Yoshikatsu Nakajima and Hideo Saito

Graduate School of Science and Technology, Keio University, Japan
{nakajima, saito}@hvrl.ics.keio.ac.jp

Abstract

Since conventional feature descriptor algorithms depend on brightness values, those algorithms cannot obtain point correspondences between aerial and map images where brightness values do not correspond but geometric structures do correspond. The research in this paper focused on obtaining point correspondences between those images. This study proposes a novel method by which curve features, edge shape, and road extraction by a CNN, independent on brightness value, produce robustness in image matching. Its effectiveness was confirmed through experiments on *Massachusetts Roads Dataset*.

Keywords: Feature descriptor, Binary features, Matching, Aerial images, Canny algorithm

1 Introduction

In recent years, due to widespread use of drones in taking aerial images, the demand for matching between aerial images and map images has increased. However, conventional feature descriptor algorithms, such as SIFT[Lowe, 2004], produce robust statistics for scaling and rotation, but do not enable one to find keypoint correspondences in images due to the changing brightness values around each keypoint. ORB[Rublee et al., 2011] and AKAZE[Alcantarilla and Solutions, 2011], conventional methods for describing features, shorten the processing time required for distance calculation during keypoint matching by describing features in binary form, however, these methods also depend on brightness values.

In this study, we propose a novel method by which curve features, edge shape, and road extraction by a CNN, independent on brightness value, produce robustness in image matching. Specifically, after extracting roads in an aerial image using a method proposed by Saito et al.[Saito et al., 2016], we detect its edge by using a Canny algorithm. Thereafter, we describe binary features for each point on the detected edge by using information as to whether or not the detected edge passes through each part of the ring-shaped feature descriptor (See Figure 1). Finally, we obtain robustness to rotation by setting the edge's orientation as the sum of vectors from the center of the ring to each part through which the edge passes.

2 Proposed Method

Figure 1 shows a flow of this proposed method. In this section, we describe the details.

2.1 Road Extraction

We employ a method proposed by Saito et al.[Saito et al., 2016], which segments an aerial image into road, building, and background classes by a CNN as pre-processing for describing features (See Figure 1, Road Extraction part). The output consists of gradient probabilities of the three classes. However, since this method describes features with curve shapes, we discard the building and background classes which have few features as a curve and process only the road class.

 ISBN 978-0-9934207-2-6

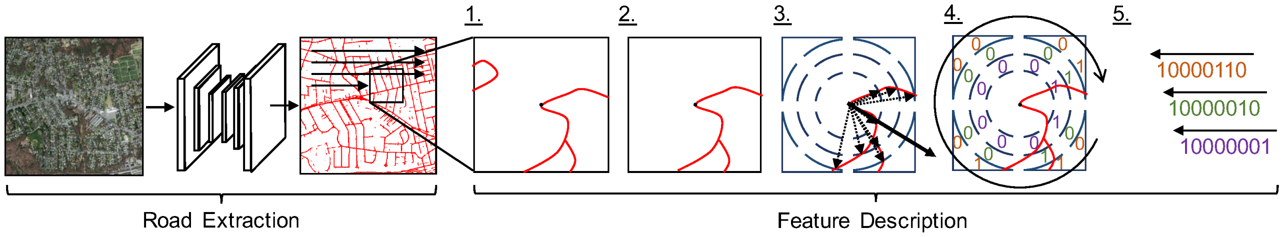

Figure 1: Flow of describing features with proposed method

2.2 Feature Desctiption

In this section, we describe the feature descriptions, which are the core of this research.

2.2.1 Generating Feature Descriptor

During pre-processing, we generate a feature descriptor $\mathscr{R}$ composed of several circles with the same center and different radii, as shown by the blue lines in Figure 1. A feature detector $\mathscr{R}$ consists of circles whose minimum radius is M and each radius is $M + k\sigma_{step}, k \subset \mathbb{Z}_{\geq 0}$. It is defined as a function which returns two values, as shown in the following equation (1) with $\boldsymbol{u} = (i, j)^{\mathrm{T}}, \boldsymbol{u} \subset \mathbb{R}^2$.

$$\mathscr{R}(\boldsymbol{u}) = (r(\boldsymbol{u}),\ a(\boldsymbol{u})) = \begin{cases} \left(\frac{\sqrt{i^2+j^2}-M}{\sigma_{step}} + 1,\ \lfloor \frac{\arctan(j,i)}{\theta} \rfloor + 1 \right) & \sqrt{i^2+j^2} \geq M \wedge \sqrt{i^2+j^2} \equiv M \pmod{\sigma_{step}} \\ (\phi,\ \phi) & otherwise \end{cases} \tag{1}$$

r indicates the number of circles to which $\boldsymbol{u}$ belongs counted from the circle with the smallest radius. a indicates the number of arcs to which $\boldsymbol{u}$ belongs counted from the top right arc when each circle is divided into arcs by the angle θ. The angle θ satisfies $l\theta = 360$ and $l \subset \mathbb{N}$. $\mathscr{R}(\boldsymbol{u})$ returns $(r(\boldsymbol{u}) = \phi, a(\boldsymbol{u}) = \phi)$ if $\boldsymbol{u}$ does not belong to any circle. Therefore, the range of r and a are $1 \leq r \leq (N - M)/\sigma_{step} + 1$ and $1 \leq a \leq 360/\theta$ except for ϕ, respectively, when the maximum radius in $\mathscr{R}$ is N. By dividing the circle by an angle, as the width of the arc becomes larger going away from the center, it is expected that robustness to a little projection distortion can be obtained in the process of describing features.

2.2.2 Detecting Edges

In this method, an edge image $\mathscr{E}(\boldsymbol{v})$ is generated by applying the Canny edge detector to the image $\mathscr{I}$. Here, the image $\mathscr{I}$ is the probability gradient of the road class as related to an aerial image, and it is a raw map as related to a map image. $\mathscr{E}(\boldsymbol{v})$ returns 1 if $\boldsymbol{v}$ is on an edge and 0 otherwise, where $\boldsymbol{v} = (x, y)^{\mathrm{T}}, \boldsymbol{v} \subset \mathbb{Z}^2$. We employed the Canny edge detector, because it is more robust to noise as compared to other edge detection methods. It is comprised of four procedures: Gauss smoothing, edge detection by the Sobel method, non-maximum value suppression, and hysteresis thresholding. For each $\boldsymbol{v}_i = (x_i, y_i)^{\mathrm{T}}$ satisfying $\mathscr{E}(\boldsymbol{v}_i) = 1$, we obtain $\mathscr{E}^c_{\boldsymbol{v}_i}$ by recursively processing eight neighbor search with $\boldsymbol{v}_i$ as the starting point after cropping $\mathscr{E}$ in the range of $x_i - N \leq x \leq x_i + N, y_i - N \leq y \leq y_i + N$, where $\boldsymbol{v}_i$ is the origin (See Figure 1, 1. and 2.). Note that $\mathscr{E}^c_{\boldsymbol{v}_i}$ has its origin at the center and has the same domain as $\mathscr{R}$.

2.2.3 Determining Orientation

Conventional methods of describing features, including SIFT[Lowe, 2004], are robust to rotation by computing orientation of each keypoint using intensity gradients around the keypoint. Even in this proposed method, the orientation of each point is determined in order to obtain robustness to rotation, however, intensity gradients are not used. Figure 1, part 3., shows a conceptual diagram on the determination of orientation. The orientation

ISBN 978-0-9934207-2-6

$\boldsymbol{o}$ at $\mathscr{E}(\boldsymbol{v}_i)$ is determined with the following equation (2) by using $\mathscr{C}_{\boldsymbol{v}_i}$, which is a set of intersection of an edge and an arc.

$$\boldsymbol{o} = \begin{pmatrix} o_x \\ o_y \end{pmatrix} = N \cdot \frac{\sum_{\boldsymbol{w} \subset \mathscr{C}_{v_i}} \boldsymbol{w}}{\|\sum_{\boldsymbol{w} \subset \mathscr{C}_{v_i}} \boldsymbol{w}\|},\ \boldsymbol{o} \subset \mathbb{R}^2,\ \mathscr{C}_{\boldsymbol{v}_i} = \left\{\boldsymbol{w} \subset \mathbb{Z}^2 | \mathscr{R}(\boldsymbol{w}) \neq (\phi, \phi) \wedge \mathscr{E}^c_{\boldsymbol{v}_i}(\boldsymbol{w}) = 1\right\} \tag{2}$$

Equation (2) shows that the vectors starting from the center of the circles and ending at the intersection of the edge and the arc are summed up and normalized the norm to N. Therefore, $\boldsymbol{o}$ calculated by the equation (2) satisfies the following equation (3).

$$\mathscr{R}(\boldsymbol{o}) = (r(\boldsymbol{o}) = \frac{N-M}{\sigma_{step}} + 1,\ a(\boldsymbol{o}) = a_o) \tag{3}$$

Since the results of the Canny edge detector include various elements, such as straight lines and isolated points, curves cannot always be easily detected. Elements other than a curve are difficult to differentiate when describing features based on edge information. To solve the problem, we do not describe features for the point $\boldsymbol{v}_i$ which satisfies $\|\sum_{\boldsymbol{w} \subset \mathscr{C}_{v_i}} \boldsymbol{w}\| < \delta_{norm}$ for threshold δ_{norm}. If the edge around $\boldsymbol{v}_i$ is an element that does not have sufficient features, such as a straight line or an isolated point, the norm of the vector indicating the orientation is considered to be sufficiently small, so that elements not having those sufficient features are excluded from processing targets.

2.2.4 Describing Binary Features

In this study, in order to shorten the processing time required for distance calculation during keypoint matching, we describe features of each point $\boldsymbol{v}_i$ in binary form. First, the binary test τ is defined as the following equation (4) using a set $\mathscr{U}_{r_i,a_j}$.

$$\tau_{\boldsymbol{v}_i}(r,\ a) = \begin{cases} 1 & \exists \boldsymbol{u} \subset \mathscr{U}_{r_i,a_j},\ \mathscr{E}^c_{\boldsymbol{v}_i}(\boldsymbol{u}) = 1 \\ 0 & otherwise \end{cases},\ \mathscr{U}_{r_i,a_j} = \left\{\boldsymbol{u} \subset \mathbb{Z}^2 | \mathscr{R}(\boldsymbol{u}) = (r_i,\ a_j)\right\} \tag{4}$$

The binary test τ returns the value of 1 or 0 by determining whether or not the edge intersects with the arc (r_i, a_j) (See Figure 1, 4.). Finally, binary features of a point $\boldsymbol{v}_i$ is described with a function f defined by the following equation (5).

$$f(\boldsymbol{v}_i) = \sum_{1 \le r \le \frac{N-M}{\sigma_{step}}+1} 2^{(r-1)\cdot\frac{360}{\theta}} g(\boldsymbol{v}_i,\ r),\ g(\boldsymbol{v}_i,\ k) = \sum_{1 \le a \le \frac{360}{\theta}} 2^{a-1} \tau_{\boldsymbol{v}_i}(k,\ a_o + a - 1 \pmod{\frac{360}{\theta}})) \tag{5}$$

After describing binary features of each circle in a feature descriptor $\mathscr{R}$, in binary form, using binary test $\tau_{\boldsymbol{v}_i}$, and with function g, we describe binary features of a point $\boldsymbol{v}_i$ by combining features of each circle with function f (See Figure 1, part 4. and 5.).

3 Experiments

In this experiment, *Massachusetts Roads Dataset* [Mnih, 2013] proposed by Mnih was used for aerial images. We captured map images corresponding to the dataset from Google Maps. The image size was aligned to 300 x 300 for both aerial images and map images. Each parameter in this experiment was $\theta = 10,\ M = 4,\ N = 30,\ \sigma_{step} = 1,\ \sigma_{norm} = 5.2$. As for the parameters of the Canny edge detector, we set the lower limit as 30 and the upper limit as 60. We conducted experiments under the following environments: CPU: Intel Core i7-6950X, GPU: GeForse GTX1080, and RAM: 128GB. After applying RANSAC to correspondences obtained by brute force matching based on hamming distance of binary features, we drew point correspondences as matching results.

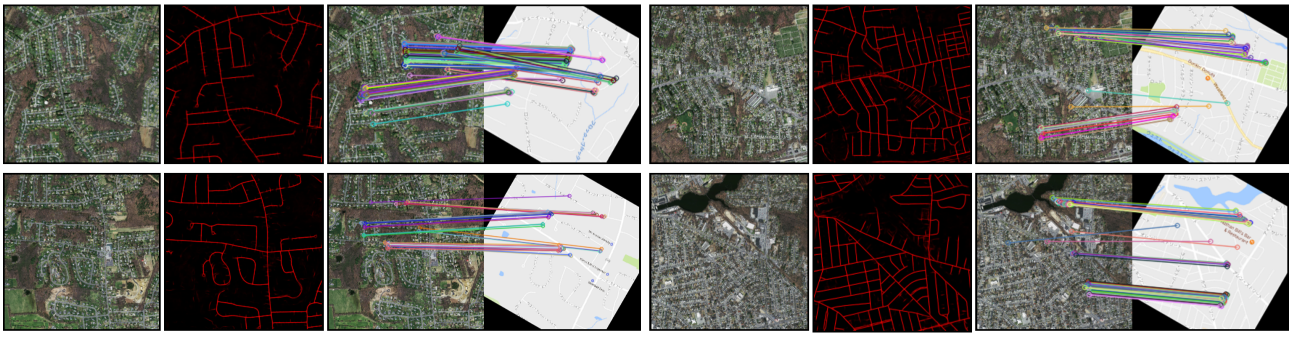

Figure 2: Example of matching results between aerial image and map (left to right: input aerial image, road extraction result, matching result)

3.1 Results

Figure 2 shows some experimental results of matching between aerial images and map images with this proposed method. We rotated map images to verify the robustness to rotation of this proposed method. Since the mean value of the number of inliers was 61.8, we were able to obtain sufficient point correspondences. Figure 2 also shows point correspondences were not biased on part of the image, but acquired from the whole. The processing time required for road extraction was 2.76 seconds on average, and the processing time required for the feature description was 0.58 seconds on average per image. Furthermore, the processing time required for matching features was 0.07 seconds on average. This is because the processing cost of distance calculation between features was reduced by describing features in binary form and using hamming distance.

4 Conclusion

In this study, we proposed a novel method by which curve features, edge shape, and road extraction by a CNN, independent on brightness value, produced robustness in image matching. We have experimentally confirmed its effectiveness and responsiveness. Our future research will focus on obtaining robustness to scaling.
Acknowledgment This research presentation is supported in part by a research assistantship of a Grant-in-Aid to the Program for Leading Graduate School for "Science for Development of Super Mature Society" from MEXT in Japan.

References

[Alcantarilla and Solutions, 2011] Alcantarilla, P. F. and Solutions, T. (2011). Fast explicit diffusion for accelerated features in nonlinear scale spaces. *IEEE Trans. Patt. Anal. Mach. Intell*, 34(7):1281–1298.

[Lowe, 2004] Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110.

[Mnih, 2013] Mnih, V. (2013). *Machine Learning for Aerial Image Labeling*. PhD thesis, University of Toronto.

[Rublee et al., 2011] Rublee, E., Rabaud, V., Konolige, K., and Bradski, G. (2011). Orb: An efficient alternative to sift or surf. In *IEEE International Conference on Computer Vision (ICCV)*, pages 2564–2571.

[Saito et al., 2016] Saito, S., Yamashita, T., and Aoki, Y. (2016). Multiple object extraction from aerial imagery with convolutional neural networks. *Electronic Imaging*, 2016(10):1–9.