Edgel Templates for Fast Planar Object Detection and Pose Estimation

Taehee Lee[*] Stefano Soatto[†]

Computer Science Department
University of California, Los Angeles

Abstract

We describe a method to select edgels and to calculate gradient orientation-based template descriptors for edgel features. An edgel is selected within a grid block based on gradient magnitude; its position and orientation are used to determine a canonical frame where the descriptor is computed based on quantized orientation. The resulting descriptor is efficiently matched using logical operations. We demonstrate the use of the resulting edgel detection and description method for planar object detection and pose estimation.

Index Terms: I.4.7 [Computing Methodologies]: Image Processing and Computer Vision—Feature Measurement; I.4.8 [Computing Methodologies]: Image Processing and Computer Vision—Scene Analysis

1 Introduction

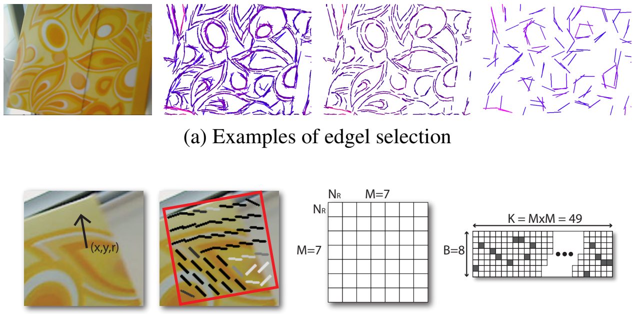
Object detection and pose estimation are important steps in registering the appearance of virtual objects in real imagery, a cornerstone of Augmented Reality (AR) applications. While "corners" are commonly used in feature tracking approaches [6, 7], there are objects and scenes with few if any distinct corners. Examples include the palm of a hand, uniformly colored objects such as walls and desk tops, line drawings. For such "textureless" objects, occluding boundaries are often the most salient and photometrically stable feature.

Taylor et al. [6] detect FAST corner features [5] from multiple synthesized viewpoints and bin the normalized intensity of the resulting (warped) images to construct a descriptor. They introduce an efficient method of comparing the binarized histograms using SIMD instructions to achieve high frame-rate and detect multiple objects in real-time. While SIFT [4] descriptors are widely used in wide-baseline matching and object detection, their computational cost hinders efficient operation on mobile devices with limited computing power, although Wagner et al. [7] have optimized SIFT descriptors for real-time detection and tracking. They also use [5] for detecting keypoints ("corners").

On the other hand, Hinterstoisser et al. [3] use dominant orientation templates to detect textureless objects. They compare orientations on regions with high gradient magnitude using SIMD operations to compute the matching score, similar to [6]. Recently, Hagbi et al. [2] introduced a shape descriptor for pose estimation for mobile augmented reality. They allow natural shapes to be a reference object, which can be hand-drawn shapes.

2 Methodology

We introduce a template based on edge segments ("edge elements", or "edgels") that describes the local appearance around a collection of edgels. Efficiently matching these template descriptors allows object detection and pose estimation under viewpoint changes and partial occlusions.



(a) Examples of edgel selection

(b) Illustration of an edgel template

Figure 1: (a) Given an image, edgels selected at multiple scales are shown. (b) For an edgel, the support region is canonized with the edgel's position and orientation. The edgel template is constructed as a matrix $\phi \in \{0,1\}^{B\times K}$

2.1 Edgel Selection

Given a grayscale image $I(x,y)$, we build an image pyramid by blurring and downsampling.For each scale-dependent sampling of the image $I(x,y;\sigma)$, at each scale level σ, we compute the first-order approximation of the image gradient $\nabla I(x,y;\sigma)$ using a 5×5 Gaussian kernel, and Sobel operators along the x and y directions. Then we divide the image into $N_B \times N_B$ grid block regions, and select the pixels that have maximum gradient magnitude in each block. Furthermore, we only select an edge pixel whose magnitude is larger than a threshold θ_1. The edge detection procedure provides a (scale-dependent) measure of edge orientation, or its normal direction r. Therefore, we represent an edge element, or "edgel", as (x,y,r,σ) with its (x,y) position, gradient orientation r, and scale σ. In Figure 1a, representative examples of selected edgels are shown.

2.2 Edgel Templates

Edgel detection provides a similarity reference frame, consisting of an origin (x,y), a direction r, and a unit σ. By assigning each of them to the same "canonical" reference, for instance via $(x,y)\mapsto(0,0)$, $r\mapsto e_1 \doteq [1,\ 0]$, and $\sigma\mapsto 1$, we obtain a description of the image that is invariant to similarity transformations locally in a neighborhood of size σ around the point (x,y).

To calculate the canonized descriptor, we sample gradient orientations $\nabla I(x,y;\sigma)/\|\nabla I\|$ in the canonical frame, that is organized into a "support region" consisting of $K = M\times M$ subregions, where each subregion is comprised of $N_R \times N_R$ pixels. Thus, a support region of an edgel (x,y,r,σ) covers $MN_R \times MN_R$ pixels centered at (x,y) and rotated so that r corresponds to the abscissa of the local reference frame. In Figure 1b, a canonized support region of an edgel is illustrated.

For each subregion, we select an edgel whose gradient magnitude is the largest within the subregion and is larger than a threshold θ_2; we then take the orientation of the selected edgel for the subregion. This is done similarly to the edgel selection in Section 2.1, but with a different threshold. Here we choose a threshold θ_2 to be smaller than θ_1 in order to make the edgel template descriptor rich

[*] e-mail: taehee@cs.ucla.edu

[†] e-mail: soatto@cs.ucla.edu

enough to describe the support region of the edgel. In our implementation, we chose $\theta_1 = 50$, and $\theta_2 = 2$.

We quantize the orientations in the edgel templates using B bins to uniformly divide $r \in [0,\ldots,\pi)$, and represent the subregion as a vector $R(r) \in \{0,1\}^B$ whose i-th element $R_i(r)$ is defined as below:

$$R_i(r) = \begin{cases} 1 & \text{if } \frac{(i-1)\pi}{B} \le r < \frac{i\pi}{B} \\ 0 & \text{otherwise} \end{cases} \quad \text{for } i = 1,2,\ldots,B \qquad (1)$$

An edgel template ϕ is finally constructed by stacking $R(r)$ vectors of K subregions as a matrix $\phi \in \{0,1\}^{B\times K}$. Note that some subregions may not have edgels with magnitudes larger than θ_2, in which case their $R(r)$ vectors are zeros. In Figure 1b, the construction of an edgel template is illustrated.

2.3 Matching Edgel Templates

The matching score of two edgel templates is computed as:

$$F(\phi_1,\phi_2) \doteq \frac{1}{K}\|\phi_1 \circ \phi_2\| \qquad (2)$$

where $\phi_1,\phi_2 \in \{0,1\}^{B\times K}$ are edgel templates, and the norm is the sum of entry-wise product of the two matrices. Hence, $F(\phi_1,\phi_2)$ is the ratio of the number of subregions in the support region that have same quantized edgel orientations between the two templates.This computation can be efficiently implemented using bit-wise logical operations and bit-count operation, similar to [6, 3].

We perform object detection and pose estimation using a coarse-to-fine matching scheme. For a test image, edgels are selected in multiple scales, and their edgel template descriptors ϕ_{test} are compared to the reference object's edgel templates ϕ_{ref}; among all putative matches with $F(\phi_{ref},\phi_{test}) > \theta_F$, we choose the matches with highest matching scores. Geometric constraints are then applied, i.e. homography for planar objects with RANSAC [1].

3 Experiments

We implemented edgel selection and template descriptor described in previous sections, and used the edgel templates for real-time object detection and pose estimation. For image sequences taken from a webcam at 640 × 480 resolution, the first frame is used as a reference object and the subsequent ones are used to detect edgels and matching their corresponding edgel templates. Then the homography is computed to render the reference object in each test image. The experiments are performed on a laptop with a 2.53GHz Intel Core 2 Duo CPU. Table 1 shows the computation time for the tasks involved in the experiments.

During these experiments, the edgel templates of the reference object are selected without synthesizing the reference image in perspective viewpoints, unlike to [6]. However, the results show that moderate to significant viewpoint changes are handled well, without time-consuming viewpoint synthesis steps for learning a reference object, as shown in Figure 2. More significant scale changes and perspective distortions can be covered when we utilize such learning procedures with edgel templates. Some failure cases can be seen in Figure 3.

Tasks	Time (ms)
grayscale image pyramids	2.7
image gradients	7.1
extracting edgel templates	16.5
matching edgel templates	20.2
homography / RANSAC	9.1
total	55.6

Table 1: *Runtime computation time.*

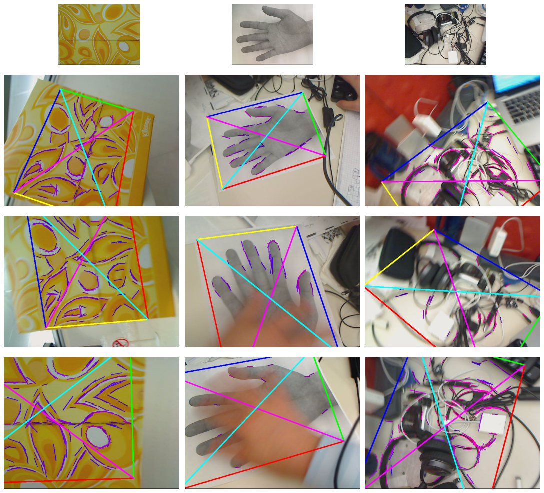


Figure 2: Representative snapshots of object detection and pose estimation under viewpoint changes and partial occlusions.

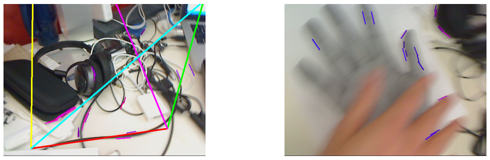


Figure 3: Failure cases: (left) incorrectly estimated homography, (right) missed target under severe blur and occlusion.

4 Conclusion

We described a method to select edgels and to calculate orientation-based edgel template descriptors. We also demonstrated using edgel templates for object detection and pose estimation under several nuisances including translation, rotation, scale changes, and occlusions. Exhaustive comparisons with different types of features and methods are left for future work. In addition, we plan to design a hierarchical selection and matching scheme for edgel templates. We also expect to optimize the implementation on mobile devices.

Acknowledgements

T. Lee was supported by Qualcomm Innovation Fellowship; Part of this research was supported by ARO 56765-CI and AFOSR FA9550-09-1-0427.

References

[1] M. A. Fischler and R. C. Bolles. Random Sample Consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 24(6):381–395, 1981.

[2] N. Hagbi, O. Bergig, J. El-Sana, and M. Billinghurst. Shape recognition and pose estimation for mobile augmented reality. In *Proc. IEEE ISMAR*, pages 65–71, 2009.

[3] S. Hinterstoisser, V. Lepetit, S. Ilic, P. Fua, and N. Navab. Dominant orientation templates for real-time detection of texture-less objects. In *Proc. IEEE CVPR*, 2010.

[4] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 2(60):91–110, 2004.

[5] E. Rosten and T. Drummond. Machine learning for high-speed corner detection. In *Proc. ECCV*, volume 1, pages 430–443, May 2006.

[6] S. Taylor and T. Drummond. Multiple target localisation at over 100 fps. In *Proc. BMVC*, September 2009.

[7] D. Wagner, G. Reitmayr, A. Mulloni, T. Drummond, and D. Schmalstieg. Real-time detection and tracking for augmented reality on mobile phones. *IEEE TVCG*, 16(3):355–368, 2010.