The entire code repository can be found at https://github.com/dgsaf/comp3007_assignment.

Contents

List of Figures

List of Tables

1   Introduction

2   Problem Analysis

We are concerned with the detection and classification of two similar but distinct types of signs. The properties specific to each type of signage are discussed in subsection 2.1 and subsection 2.2. Here, we remark on the properties common to both of them, which are numerous and allow for a broadly unified approach to their detection and classification.

Each sign consists of a black background with white characters (digits and directional arrows) forming the foreground of the sign. The black background is rectangular in shape, and bounds the characters regions. Furthermore, the contrast between the black background and the white foreground is strong across all colour channels and in grayscale.

The digits are uniform in their construction, being printed in a monospace font with fixed height and allocated a fixed width; although each digit does not necessarily fully extend across its allocated width - as can be seen for the digit 1. Likewise, the directional arrows are also uniform in their construction but their dimensions are distinct from those of the digits - noticeably having a smaller height than the digits.

The digits do not overlap with each other (unless subject to camera artifacts/blurring) but are closely adjacent to each other, and regularly spaced. Each sign contains at least one sequence of three digits, which are ideally surrounded by a black sub-region of the sign; although for task 2, the gap between the leftmost digit and the edge of the sign can be very small.

In summary, the following properties of the signs are almost always observed, providing a robust foundation for their detection:

- The uniformly white colour of the characters;

- The uniformly black background of the sign;

- The strong contrast between the white characters and the black background across all red, green, and blue channels as well as in grayscale;

- The uniform height of the digits;

- The uniform spacing of the characters horizontally;

- The height of the digits exceeding the width;

- The uniform dimensions of the arrows;

- The presence of a chain of exactly three digits;

- A sizeable sub-region of the sign above and below the chain of digits, which is uniformly black.

However, the assumption of the above properties may be invalidated by any one of the (non-exhaustive) list of conditions:

- The presence of strong radial, or motion blurring which may cause character to overlap;

- The presence of strong shadowns, glare, or other variations in lighting conditions across the character;

2

- The presence of a foreign object, such as a sticker or a mark, on the sign (violating the assumptions on uniformity of layout) or across the characters (violating the assumptions on uniformity of characters).

It should also be noted that often there are other monospace white characters present near the sign - typically letters forming the name of the location marked by the sign. These characters can be distinguished from the digits (and arrows) by the fact that they almost always form chains of more than three characters, and presented on their own signs which almost never have a black background.

2.1   Task 1

Further refinement of the problem analysis is possible for task 1. We are concerned with finding only one chain of three digits, and we are not concerned with directional arrows at all. Futhermore, there exists a sizeable gap between the chain of digits and the edge of the sign both horizontally and vertically.
Bricks and other objects which are of approximately uniform dimensions and uniformly spaced are regularly present in images for task 1 - hence, when exploiting the uniformity of the layout of the digits, care must be taken to avoid or handle the false detection of these other objects.

Regular Cases

Outlier Cases

2.2   Task 2

Regular Cases

Outlier Cases

3   Implementation

3.1   Overview

To detect the characters (digits and arrows), we utilise the technique of Maximally Stable Extremal Regions (MSER). We suggest that this technique is suitable for our purpose, as the characters are uniformly white and contrast well against the surrounding uniformly black area of the signs. Hence, these characters should form regions which are very stable under thresholding, when compared with their local context. Furthermore, this technique should be fairly robust to the presence of shadows and other variations in illumination.

3.2   Mathematical Details

Image

We define an image I, with n channels and of width w and height h, to be a mapping

$$I : D \to S : (x, y) \mapsto I(x, y) = \big(I_1(x, y), \ldots, I_n(x, y)\big), \tag{1}$$

where the domain $D \subset \mathbb{N}^2$ is of the form $D = \{0, \ldots, w - 1\} \times \{0, \ldots, h - 1\}$, and S is the codomain. Constraints can be specified on S, but for our purposes it suffices to suppose that either $S \subseteq \{0, \ldots, 255\}^n$ or $S \subset \mathbb{R}^n$.

Thresholded Image

Suppose $I : D \to S \subset \mathbb{R}$ is a single-channel image. For each $t \in S$, we define the Boolean-valued image I_t to be of the form

$$I_t : D \to \mathbb{B} : (x, y) \mapsto \begin{cases} 0 & \text{if} \quad I(x, y) \leq t \\ 1 & \text{if} \quad I(x, y) > t \end{cases}, \tag{2}$$

and we say that I_t is a thresholded image.

Bounding Box

Suppose D is the domain of an image $I : D \to S$. For all $R \subseteq D$, we define the bounding box $B(R)$ of R to be

$$B(R) = [x_R, x_R + w_R] \times [y_R, y_R + h_R] \quad \text{such that} \quad R \subseteq B(R) \tag{3}$$

and where $w_R, h_R \in \mathbb{N}$ are minimal.

Connectedness

Suppose D is the domain of an image. An adjacency relation A on D is a Boolean-valued mapping

$$A : D \times D \to \mathbb{B} : (p, q) \mapsto A(p, q), \tag{4}$$

which indicates if the two points of the domain are considered to be adjacent. For any $p \in D$, we may define the neighbourhood $N(p)$ of p to be the set of all points which are adjacent to it; that is,

$$N(p) = \{q \in D \mid A(p, q)\}. \tag{5}$$

For any $p, q \in D$, we say that p and q are connected if there exists a finite sequence $(\rho_k)_{1 \leq k \leq n}$ in D such that

$$A(p, \rho_1) \wedge A(\rho_1, \rho_2) \wedge \cdots \wedge A(\rho_{n-1}, \rho_n) \wedge A(\rho_n, q) = 1. \tag{6}$$

In the case where the adjacency relation A is symmetric, then connectedness defines an equivalency relation; whence we may write, for any connected $p, q \in D$ that $p \sim q$.

Note that we are primarily concerned with the adjacency relations associated with the Von Neumann neighbourhood (4-connectivity)

$$N_4(p) = \{p + n \in D \mid n \in \{(0, 1), (1, 0), (0, -1), (-1, 0)\}\}, \tag{7}$$

and the Moore neighbourhood (8-connectivity)

$$N_8(p) = \{p + n \in D \mid n \in \{-1, 0, 1\} \times \{-1, 0, 1\} \setminus (0, 0)\}. \tag{8}$$

Region

Suppose D is the domain of an image $I : D \to S$ and let $A : D \times D \to \mathbb{B}$ be a symmetric adjacency relation on D. We say that $R \subseteq D$ is a region if every element of R is connected to every other element of R; that is,

$$p, q \in R \implies p \sim q. \tag{9}$$

We define the (inner) boundary ∂R of a region R to be subset of points of R which are also connected to at least one point not in R; that is,

$$\partial R = \{p \in R \mid \exists q \in D \setminus R : A(p, q)\}. \tag{10}$$

We define the outer boundary ΔR of a region R to be the set of points of D which do not belong to R but are adjacent to a point of R; that is,

$$\Delta R = \{p \in D \setminus R \mid \exists q \in R : A(p, q)\}. \tag{11}$$

Maximally Stable Extremal Region (MSER)

Suppose $I : D \to S \subset \mathbb{R}$ is a single-channel image, and suppose $A : D \times D \to \mathbb{B}$ is the (symmetric) adjacency relation associated with either the Von Neumann neighbourhood or the Moore neighbourhood. Suppose $R \subseteq D$ is a region. We say that R is a minimal region if for all $p \in R$ and $q \in \Delta R$ we have $I(p) < I(q)$; which is equivalently written as the requirement that

$$\max_{p \in R} I(p) < \min_{q \in \Delta R} I(q). \tag{12}$$

Similarly, we say that R is a maximal region if for all $p \in R$ and $q \in \Delta R$ we have $I(p) > I(q)$; which is equivalently written as the requirement that

$$\min_{p \in R} I(p) > \max_{q \in \Delta R} I(q). \tag{13}$$

We say that R is an extremal region if it is either a minimal or maximal region.
The formulation of extremal regions in terms of the minimal and maximal intensity values of the image permits the usage of thresholding. Suppose that R is an extremal region, and suppose that $t \in S$. Consider the thresholded region R_t, defined by

$$R_t = \{p \in R \mid I(p) < t\}, \tag{14}$$

which is itself an extremal region, and for which we have that

$$\max_{p \in R_t} I(p) < t. \tag{15}$$

We note that $R_t \subseteq R$ for all $t \in S$. We also note that when $t_1 \leq t_2$, we have that $R_{t_1} \subseteq R_{t_2}$; that is, the thresholded regions form an increasing (by set inclusion) sequence of subsets of R. For any increasing chain $t_1 < \ldots < t_n$ in S, we have

$$\emptyset \subseteq R_{t_1} \subseteq \cdots \subseteq R_{t_n} \subseteq R. \tag{16}$$

In the MSER approach, the stability of an extremal region R is measured by examining the change in the cardinality of R_t with the change in the threshold t. That is, for a particular threshold $t \in S$

and threshold step $\delta \in S$, such that $t - \delta, t + \delta \in S$, the rate of growth of the extremal region R is given by

$$G_\delta(R;t) = \frac{|R_{t+\delta} \setminus R_{t-\delta}|}{|R_t|}. \tag{17}$$

An extremal region R_{t_0} is then said to be maximally stable if $G_\delta(R;t)$ has a local minimum at $t = t_0$. Such a thresholded region experiences minimal change (in cardinality) when the threshold t_0 is increased/decreased by δ.

3.3 Task 1

Detection

Classification

3.4 Task 2

Detection

Classification

4 Validation Performance

4.1 Task 1

4.2 Task 2

5 Conclusion

A   Source Code