Defining a Metric for the Space of Images

M204 Home Work

Roshan Mathew Philip
2211192

January 2024

1 Introduction

When do we say two images are similar to each other? This is the question that we aim to answer by defining a metric on the set of all images. If the distance between two images is small(smaller than a number we decide), naturally the images are similar.

In this work, we consider images to be $m \times n$ matrices with entries from $\{0, 1\}$. This is reminicent of the displays used in digital monitors, with the entries corresponding to a pixel on the display. Making the matrix entries from $\{0, 1\}$ unfortunately means that our images will not have colours or even different brightness levels. Towards the end of this work we will present ideas to potentially overcome this limitation. Throughout this work we will refer to the matrix entries with value 1 as "on" or "bright" pixels, and matrix entries with value 0 as "off" or "dark" pixels.

2 Taking care of translation between the images

Consider the two images given below (* represents a 1, $\cdot$ represents a 0).

$$\begin{pmatrix} \cdot & \cdot & \cdot & \cdot \\ \cdot & * & * & \cdot \\ \cdot & * & * & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{pmatrix} \qquad \begin{pmatrix} \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & * & * \\ \cdot & \cdot & * & * \\ \cdot & \cdot & \cdot & \cdot \end{pmatrix}$$

(a) Image 1 (b) Image 2

We can clearly see that these are the same images, just shifted over. To quantify the shift, we just take the distance between the average positions of bright pixels in each image.
The positions of the bright pixels in image 1 are $(2, 2), (2, 3), (3, 2), (3, 3)$, and the average is (2.5,2.5).
Similarly the average position of the bright pixels in image 2 is $(2.5, 3.5)$.
We then use the Euclidean metric (L^2-metric) for $\mathbb{R}^2$ to calculate the distance between these points. We take this as the distance by which the images are shifted from each other. We notate this number as

$$d_T(\text{image1, image2})$$

This constitutes one part of our metric.

2.1 Scaling differences between the images

Consider the two images given below.

$$\begin{pmatrix} * & * & * & * & * \\ * & \cdot & \cdot & \cdot & * \\ * & \cdot & \cdot & \cdot & * \\ * & \cdot & \cdot & \cdot & * \\ * & * & * & * & * \end{pmatrix} \qquad \begin{pmatrix} \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & * & * & * & \cdot \\ \cdot & * & \cdot & * & \cdot \\ \cdot & * & * & * & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \end{pmatrix}$$

(a) Image 1 (b) Image 2

These are clearly the same image just differing in size. We use the difference in standard deviation of the position of the bright pixels from its average position as a measure of the difference in size between the images.

$$\sigma_{\text{img1}} = \sqrt{\frac{1}{N} \sum \left(L^2((i,j), \overline{(i,j)}) \right)^2}$$

where N is the total number of bright pixels in img1, the summation is over all the bright pixles in img1, and $\overline{(i,j)}$ is the mean position of all the bright pixels.

$$d_\sigma = |\sigma_{\text{img1}} - \sigma_{\text{img2}}|$$

This constitutes the second part of our metric.

3 Chamfer distance

Consider an image $A \in M_{m\times n}(\{0,1\})$. We construct a new set from A

$$\widetilde{A} = \left\{ \frac{1}{\sigma_A}\left[(i,j) - \overline{(i,j)}\right] \middle| A_{ij} = 1 \right\}$$

We call $\widetilde{A}$ the normalised image of A. The intuition is that the translation and scaling of the image have been taken care of by taking the distance vector from the mean point and scaling it with $1/\sigma$.

Consider two normalised images $\widetilde{A}, \widetilde{B}$.
We define d^A_{ij} as the L^2-distance from (i,j) to the closest point in $\widetilde{A}$. Put a bit more symbolically,

$$d^A_{ij} := \min\{ L^2\left((i,j),(p,q)\right) \big| (p,q) \in \widetilde{A}\}$$

We now define the chamfer distance as:

$$d_C(\widetilde{A},\widetilde{B}) := \sum_{(i,j)\in\widetilde{A}} d^B_{ij} + \sum_{(i,j)\in\widetilde{B}} d^A_{ij}$$

Claim: The chamfer distance is a well-defined metric
Proof:

- d^A_{ij} is a well defined function from (the set of normalised images)$\times$(the set of normalised images) to the non-negative reals. $\therefore$ d_C, which is a sum of these, is also well-behaved and its range is the non-negative reals.

- $$\begin{aligned} d_C(\widetilde{A},\widetilde{B}) &= \sum_{(i,j)\in\widetilde{A}} d^B_{ij} + \sum_{(i,j)\in\widetilde{B}} d^A_{ij} \\ &= \sum_{(i,j)\in\widetilde{B}} d^A_{ij} + \sum_{(i,j)\in\widetilde{A}} d^B_{ij} \\ &= d_C(\widetilde{B},\widetilde{A}) \end{aligned}$$

- Let $\widetilde{A} = \widetilde{B}$.
  For any $(i,j) \in \widetilde{A}$,
  $\min\{L^2((i,j),(p,q)) | (p,q) \in \widetilde{A}\}$
  $= L^2((i,j),(i,j)) = 0$
  $\therefore\ d^A_{ij} = 0$
  $\implies d_C(\widetilde{A},\widetilde{A}) = 0$

  Let $d_C(\widetilde{A},\widetilde{B}) = 0$
  Since d_C is the sum of non-negative reals, each term is individually zero.

  $$\begin{gathered} \forall (i,j) \in \widetilde{A}, \min\{L^2((i,j),(p,q)) | (p,q) \in \widetilde{B}\} = 0 \\ \because L^2((i,j),(p,q)) = 0 \iff (i,j) = (p,q) \\ \implies (i,j) \in \widetilde{B} \\ \therefore \widetilde{A} \subset \widetilde{B} \end{gathered}$$

  Similarly

  $$\begin{gathered} \forall (i,j) \in \widetilde{B}, d^A_{ij} = 0 \\ \implies \widetilde{B} \subset \widetilde{A} \\ \therefore \widetilde{A} = \widetilde{B} \end{gathered}$$

- Consider three normalised images $\widetilde{A}$, $\widetilde{B}$, and $\widetilde{C}$.

  Let $(i,j) \in \widetilde{A}$ and let $(u,v) \in \widetilde{B}$ be the closest point in $\widetilde{B}$ to (i,j) according to L^2

  Let $(p,q) \in \widetilde{C}$ be the closest point to (u,v) according to the L^2 norm, i.e, $d^C_{uv} = L^2((u,v),(p,q))$

  By the triangle inequality of L^2

  $$L^2((i,j),(p,q)) \le L^2((i,j),(u,v)) + L^2((u,v),(p,q))$$

  $$\begin{gathered} \because d^C_{ij} \le L^2((i,j),(p,q)) \\ d^C_{ij} \le d^B_{ij} + d^C_{uv} \\ \therefore \sum_{(i,j)\in\widetilde{A}} d^C_{ij} \le \sum_{(i,j)\in\widetilde{A}} d^B_{ij} + \sum_{(u,v)\in Q} d^C_{uv} \end{gathered}$$

  Where Q is the set of all points in B s.t it is closest under L^2 to some point in $\widetilde{A}$

  $$\implies \sum_{(i,j)\in\widetilde{A}} d^C_{ij} \le \sum_{(i,j)\in\widetilde{A}} d^B_{ij} + \sum_{(u,v)\in Q} d^C_{uv} + \sum_{(u,v)\in B-Q} d^C_{uv}$$

  $$\implies \sum_{(i,j)\in\widetilde{A}} d^C_{ij} \le \sum_{(i,j)\in\widetilde{A}} d^B_{ij} + \sum_{(u,v)\in\widetilde{B}} d^C_{uv} (3.1)$$

  Similarly

  $$\implies \sum_{(i,j)\in\widetilde{C}} d^A_{ij} \le \sum_{(i,j)\in\widetilde{C}} d^B_{ij} + \sum_{(u,v)\in\widetilde{B}} d^A_{uv} (3.2)$$

  Adding (3.1) and (3.2) we get

  $$d_C(A,C) \le d_C(A,B) + d_C(B,C)$$

Therefore the chamfer distance is a metric on the space of normalised images.

4 Taking care of rotations

We are familiar with the 2-D rotation matrix

$$R(\theta) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

with $\theta \in [0, 2\pi]$ The rotated image $R(\theta)\widetilde{A}$ is another normalised image obtained by acting on the elements of $\widetilde{A}$ by the rotation operator. Let

$$\begin{aligned} r_A &:= \inf\left\{ d_C(R(\theta)\widetilde{A}, \widetilde{B}) \right\} \\ r_B &:= \inf\left\{ d_C(\widetilde{A}, R(\theta)\widetilde{B}) \right\} \\ r &:= \min(r_A, r_B) \\ d_\theta(\widetilde{A}, \widetilde{B}) &:= \text{minimum angle } \theta \text{ at which} \\ &\qquad\quad r \text{ attains its infimum value.} \end{aligned}$$

We define the rotation-inclusive chamfer distance as the chamfer distance of the closest rotated images, i.e,

$$\begin{aligned} d_{RC} := \min\Big\{ & d_C\left\{ R(d_\theta)\widetilde{A}, \widetilde{B} \right\}, \\ & d_C\left\{ \widetilde{A}, R(d_\theta)\widetilde{B} \right\} \Big\} \end{aligned}$$

Here $d_\theta = d_\theta(\widetilde{A}, \widetilde{B})$

Claim: d_θ is a 'semi'-metric on the space of normalised images.
Proof:

- From the definition $d_\theta(\widetilde{A}, \widetilde{B}) \in [0, \pi)$
- We can see that the definition of d_θ is symmetric with respect to $\widetilde{A}$ and $\widetilde{B}$.
  $\therefore\ d_\theta(\widetilde{A}, \widetilde{B}) = d_\theta(\widetilde{B}, \widetilde{A})$
- Since $d_C(\widetilde{A}, \widetilde{A}) = 0$, we can see that $d_\theta(\widetilde{A}, \widetilde{A}) = 0$
  But the reverse implication is not true, $d_\theta(\widetilde{A}, \widetilde{B}) = 0 \;\not\Longrightarrow\; \widetilde{A} = \widetilde{B}$. It may just happen that the closest rotated images of the two are just the original unrotated images.

5 The final metric

Using the fact that any linear combination of metrics also results in a metric[1], we define the final metric on the space of images as:

$$\begin{aligned} d(A, B) := {} & \lambda_T d_T(A, B) + \lambda_\sigma d_\sigma(A, B) \\ & + \lambda_\theta d_\theta(A, B) + \lambda_{RC} d_{RC}(A, B) \end{aligned} \tag{5.1}$$

The coefficients of each of these individual metrics are weights that we can tweak depending on our requirements. For example, if we want our metric to be very sensitive to rotations, we can make the weight of the rotation distance higher.

6 Examples

Let

$$A = \begin{pmatrix} \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & * & * & * & * & * & * & * & \cdot \\ \cdot & \cdot & * & \cdot & \cdot & \cdot & \cdot & \cdot & * & \cdot \\ \cdot & \cdot & * & \cdot & \cdot & \cdot & \cdot & \cdot & * & \cdot \\ \cdot & \cdot & * & \cdot & \cdot & \cdot & \cdot & \cdot & * & \cdot \\ \cdot & \cdot & * & \cdot & \cdot & \cdot & \cdot & \cdot & * & \cdot \\ \cdot & \cdot & * & * & * & * & * & * & * & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{pmatrix}$$

$$B = \begin{pmatrix} \cdot & \cdot & \cdot & \cdot & * & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & * & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & * & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & * & \cdot & \cdot & \cdot & \cdot & \cdot \\ * & * & * & * & * & * & * & * & * & \cdot \\ \cdot & \cdot & \cdot & \cdot & * & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & * & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & * & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & * & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{pmatrix}$$

Mean position of bright pixels of $A = (3.5, 5)$
Mean position of bright pixles of $B = (4, 4)$
$\therefore\ d_T = 1.118$
Standard deviation in $A = 3.20$
Standard deviation in $B = 2.66$
$\therefore\ d_\sigma = 0.54$
$d_\theta = 2.411$
$d_{RC} = 13.06$
Assigning the weights as
$\lambda_T = 0.1, \lambda_\sigma = 0.25, \lambda_\theta = 0.5, \lambda_{RC} = 1$
We get the final metric as $d = 16.448$

7 Disadvantages of the metric

- Struggles to distinguish shapes when they are too small. For example a 3×3 square and a 3-sided triangle (in units of pixels)will be very close.
  This is seen almost everywhere else too, even the human eye fails to distinguish shapes that are too small.
- It is very computationally expensive, especially for large images. To reduce this there are various approximations that the field of computer vision uses.

[1] $d_{RC} = 0$ and $d_\theta = 0 \implies \widetilde{A} = \widetilde{B}$, this is enough to prove $d(\widetilde{A}, \widetilde{B}) = 0 \iff \widetilde{A} = \widetilde{B}$, even though this is not true for d_θ alone. For applications where we need triangle inequality to be valid, we can ignore the rotation part and define the metric using just the chamfer distance.

8 Expanding to greyscale images

To bring in the concept of 'colours' or 'shades', we let the matrix elemnts of the images range from 0 to C, where $C \in \mathbb{N}$. An image now is an $m \times n$ matrix with elements from the set of natural numbers between and including 0 and C

8.1 Edge detection

We find pixels whose values are sufficiently different from its neighbouring pixels and call it an edge pixel. The intuition is that at edges the 'colour' will change drastically, and therefore the neighbourhood of an edge pixel will have wildly different pixel values. We make a new $m \times n$ matrix with by replacing all edge pixels with value 1 and replacing everything else with value 0. We then use (5.1) to compare these 'edge' images.

References

1. Luke Hawkes, *A visual representation of the Chamfer distance function*, https://www.youtube.com/watch?v=P4IyrsWicfs