

# Όραση Υπολογιστών: Εργασία 1

Η παρούσα εργασία έχει ως στόχο την ανάλυση της δομής εγγράφων και χρησιμοποιεί δεδομένα εικόνων που περιέχουν σκαναρισμένα τυπωμένα κείμενα και βρίσκονται στον παρακάτω σύνδεσμο:

<https://vc.ee.duth.gr:6960/index.php/s/Z1HzaMKB5lUrtAD>

Στην παραπάνω συλλογή υπάρχουν 10 διαφορετικές εικόνες. Ο κάθε φοιτητής καλείται να επιλέξει μία από αυτές τις εικόνες βάσει του τελευταίου ψηφίου του ΑΕΜ του, π.χ. οι φοιτητές που το τελευταίο ψηφίο του μητρώου τους τελειώνει σε 2, επιλέγουν τη “2.png”. Εφόσον ο φοιτητής έχει επιλέξει την κατάλληλη εικόνα, στη συνέχεια θα πρέπει να δημιουργήσει δύο αντίγραφα αυτής στα οποία να προσθέσει θόρυβο, “Gaussian” στο 1<sup>ο</sup> αντίγραφο και “Salt and Pepper” στο 2<sup>ο</sup>.

Έπειτα, ο κάθε φοιτητής θα πρέπει να αναπτύξει μια ολοκληρωμένη μεθοδολογία και να την υλοποιήσει σε κώδικα, ώστε για τις δύο εικόνες (την πρωτότυπη και αυτή με το θόρυβο) να επιτύχει τα εξής:

1. **Ανίχνευση όλων των υποπεριοχών** κειμένου του εγγράφου. Ως **υποπεριοχή** ορίζεται ένα τμήμα του εγγράφου με κείμενο, το οποίο διακρίνεται ως προς τη θέση του από τα υπόλοιπα μέρη του εγγράφου (π.χ. αρίθμηση σελίδας, υποσέλιδο, παράγραφος, τίτλος κ.α.). Το πρόγραμμα θα πρέπει να εμφανίζει και να αποθηκεύει την εικόνα εγγράφου στην οποία να είναι εμφανείς οι διαφορετικές υποπεριοχές της. Συγκεκριμένα, για κάθε υποπεριοχή να σχεδιαστεί ένα **περιβάλλον κουτί** (bounding box) καθώς και ένας **μοναδικός αύξων αριθμός**. Για την επίλυση του ερωτήματος αυτού, να ΜΗΝ χρησιμοποιήσετε τη συνάρτηση ‘cv2.findContours’.
2. Για κάθε μια υποπεριοχή να μετρηθούν και να εμφανίζονται ως έξοδος του προγράμματος τα ακόλουθα μεγέθη:
  - α. Η **επιφάνεια της υποπεριοχής που καταλαμβάνεται από κείμενο**, που ορίζεται ως ο αριθμός των εικονοστοιχείων που ανήκουν σε γράμμα (σκουρόχρωμα) και όχι στη σελίδα (ανοιχτόχρωμα).
  - β. Η **επιφάνεια του περιβάλλοντος κουτιού** (bounding box) της υποπεριοχής.
  - γ. Ο **αριθμός των λέξεων** που εμπεριέχονται στην υποπεριοχή.
  - δ. Η **μέση τιμή διαβάθμισης του γκρι** των εικονοστοιχείων που περιέχονται στα περιβάλλοντα κουτιά (bounding boxes) των αντικειμένων, με τέτοιο τρόπο ώστε η **ταχύτητα εκτέλεσης υπολογισμού να είναι ανεξάρτητη του μεγέθους της υποπεριοχής**. Για την επίλυση του ερωτήματος αυτού, να ΜΗΝ χρησιμοποιήσετε τη συνάρτηση ‘cv2.integral’ αλλά δική σας υλοποίηση συνάρτησης που να κάνει χρήση της εικόνας ολοκληρώματος (integral image).

Ο κώδικας που θα παραχθεί θα πρέπει να συνοδεύεται από τεχνική αναφορά η οποία θα πρέπει να περιλαμβάνει τα παρακάτω χαρακτηριστικά:

1. Να περιγράφονται και να αναλύονται τα βήματα και οι επιμέρους μέθοδοι που απαρτίζουν την ολοκληρωμένη μεθοδολογία.
2. Να παρουσιάζονται και να σχολιάζονται τα αποτελέσματα των ενδιάμεσων βημάτων, τα οποία αιτιολογούν την επιλογή της συγκεκριμένης μεθόδου.
3. Να παρουσιάζονται τα τελικά αποτελέσματα.
4. Να δικαιολογούνται πιθανές αστοχίες της μεθοδολογίας με συγκεκριμένα χαρακτηριστικά των εικόνων που τις προκαλούν.
5. Να συγκρίνεται ποιοτικά η επίδοση της εφαρμογής στις περιπτώσεις ύπαρξης και μη θορύβου.
6. Να υλοποιηθεί και δεύτερη μέθοδος εντοπισμού των **υποπεριοχών και λέξεων** του εγγράφου με χρήση της μεθόδου **projection profile**. Τα αποτελέσματα της μεθόδου αυτής να συγκριθούν με εκείνα της μεθόδου που χρησιμοποιήσατε και να σχολιαστούν πιθανά πλεονεκτήματα, περιορισμοί ή περιπτώσεις απόκλισης μεταξύ των δύο προσεγγίσεων. Στο τέλος της εκφώνησης παρέχεται ένας αλγόριθμος για τη μέθοδο projection profile.

## ΥΠΟΔΕΙΞΗ

Για την επίτευξη των παραπάνω ζητούμενων, προτείνεται:

1. Στην αρχική εικόνα να εφαρμόσετε γραμμικό ή μη γραμμικό φίλτρο απόρριψης θορύβου της επιλογής σας. Το βήμα αυτό θα πρέπει να υλοποιηθεί **ΧΩΡΙΣ** τη χρήση των αντίστοιχων **συναρτήσεων της OpenCV**.

2. Στην εικόνα του αποτελέσματος του παραπάνω βήματος να εφαρμόσετε κατάλληλο κατώφλι για τη μετατροπή της εικόνας διαβάθμισης του γκρι σε δυαδική εικόνα, χρησιμοποιώντας τη συνάρτηση 'cv2.threshold'
3. Για να γίνουν οι ζητούμενες μετρήσεις θα πρέπει να εφαρμόσετε κατάλληλη μεθοδολογία που να βρίσκει συνδεδεμένα αντικείμενα, που υλοποιείται με τη συνάρτηση 'cv2.connectedComponents'.
4. Για την εύρεση του περιβάλλοντος κουτιού που χαρακτηρίζει μία περιοχή, μπορείτε να χρησιμοποιήσετε τη συνάρτηση 'cv2.boundingRect'.
5. Για την σχεδίαση παραλληλόγραμμων και κειμένου σε μια εικόνα, μπορείτε να χρησιμοποιήσετε τις συναρτήσεις 'cv2.rectangle' και 'cv2.putText', αντίστοιχα.

Στο παρακάτω σχήμα φαίνεται ένα υπόδειγμα της παραγόμενης εξόδου του προγράμματος προς ανάπτυξη:

**1** Blockage is not a primary issue. This has been verified experimentally in [7], where it is shown that a 9-dB improvement is found when utilizing a Bessel beam in sub-THz links compared to a Gaussian beam. Bessel beams have previously been utilized in the optical regime and are not also being explored at millimeter-wave (mmWave) and (sub-)THz frequencies. All of these properties as well as a far simpler linearly varying phase profile (as opposed to the spherical requirement in beamfocusing) make Bessel beams a promising wavefront candidate for THz communications.

**4** *Curved and Airy beams*  
While Bessel beams can go "through" an obstacle, another class of beams can completely circumvent blockage while also potentially enabling NLOS links. These beams follow a curved trajectory, and when viewed from the transverse direction, the beams appear to have acceleration without any external energy—earning their classification as *self-accelerating beams*. Such beams have been recently demonstrated [5], where the beam profile is described with the Airy function. The Airy function is a solution to the Stokes equation [5]. The cross-sectional intensity of such Airy beams is illustrated in Figure 6(d), and these beams satisfy a solution to Maxwell's equations with an exponential cubic phase front. The electric field in one dimension is given as

$$E(x, z) = A_i \left( \gamma x - \frac{\gamma^2 z^2}{4k^2} + j \frac{\alpha \gamma z}{k} \right) \exp(j\Phi(x, z) + \alpha \left[ x - \frac{\gamma^2 z^2}{2k^2} \right]) \quad (19)$$

**6** Here  $A_i(\cdot)$  is the Airy function [5]. The phase is given by  $\Phi(x, z)$ . The parameter  $\alpha$  specifies an attenuation constant to make the energy requirement of the beam possible. The beam follows a curved path  $x(z) = \gamma^2 z^2 / 4k^2$ , with the initial electric field at  $x, z = 0$  satisfying the amplitude  $A_i(\gamma x)$  and phase  $\arg(A_i(\gamma x))$ . The curvature function depends on  $x_0$ , which is the initial placement of the beam on the aperture. The trajectory can be adapted based on the properties of the environment (e.g., the blocker location and size) for resilient, high-speed connectivity in sub-THz regimes [10], [14].

The required phase and corresponding wavefront to generate these beams can be found by first describing the desired curvature trajectory. Then, the principle of caustics from ray optics can be utilized to find the required phase. That is, we may first consider an arbitrary curve (see Figure 6). Then, tangents from the curve can be drawn to the aperture, and the required phase at that point in the radiating aperture is found as  $\Phi(x)$  [14]

$$\Phi(x) = k \frac{d(x(z)/z)}{\sqrt{1 + dx(z)^2/z^2}} \quad (8)$$

**9** Here  $x(z)$  defines the parabolic desired trajectory that we wish to engineer across the beam propagation. Although shown for a 1D aperture in the  $x$ -plane, it also easily extends to an aperture in the  $xy$  plane as well, as shown in [5] (see Figure 6). The important factor is that the greater the curvature, the larger the aperture size needed to fulfill the tangential requirements. As

**2** obvious, an increased beam curvature requires both a steeper phase progression across the array aperture as well as a much larger aperture size. These beams are extremely promising to avoid blockage, and can also be utilized to curve around corners, enabling potential NLOS links. Since the angle of arrival can be changed when incident upon the receiver, these beams can also be utilized for beam alignment and link monitoring.

#### Wavefront generation

Once the appropriate beam type is selected, the corresponding wavefront can be identified by following the Huygens–Fresnel principle. However, on the one hand, beamforming antenna arrays at THz frequencies are still under development, and on the other hand, phases are discretized both spatially because of both the size of the antenna element and also the sampling limitations of the corresponding phase shifter [13]. In this context, graphene-based plasmonic antenna arrays have also been proposed [1]. It has been shown that these antennas can be up to an order of magnitude smaller than a conventional patch antenna, allowing for dense integration within an array. With the plasmonic modulator, it becomes possible to apply any phase across the radiating element, thus providing adequate support for the nonlinear phase requirements of beamfocusing and Airy beams, in addition to the linear ramp of Bessel and beamforming.

Taking this concept further still, metasurfaces can also be utilized to generate near-field beams [1], [23]. These are tightly coupled 2D counterparts of metamaterials, where a sub-wavelength radiating element—usually a metal half-ring—is utilized to produce a particular radiation pattern. With metasurfaces, the radiating element can reach the level of a wavelet—being extremely small. However, the drawback is that the entire metasurface response is configured at once, and thus it isn't straightforward to find the phases and required configuration of applying the wavefront required to generate a particular beam. Further yet, metasurface designs are heavily frequency dependent; thus, just because a design works at one frequency doesn't mean it can be scaled to another. Instead, significant reworking is required whenever the design parameters, such as the signal wavelength or bandwidth, are changed. Nonetheless, metasurfaces have great potential for use in near-field radiation, and their design is a subject of active research.

In addition to active arrays and metasurfaces can also be utilized in reflection, whereby a reconfigurable intelligent surface (RIS) can be utilized to impart the necessary phase on a beam that is incident upon it [1], [7]. At the same time, it becomes possible to design and 3D-print a specific lens that can be fitted atop a large horn lens antenna or to utilize a complicated antenna configuration, such as a radiating dish [7], [13]. Here, the resolution of the lens plays a role in how well the beam is generated. Notably, this is how the majority of experimental works are being validated since the design process is much simpler and can be completed with currently available technology. For example, the work in [7] utilized a 3D-printed lens mounted atop an 11.8-cm-wide horn lens antenna to generate an ultrabroadband Bessel beam (20-GHz bandwidth and

-----Region 1 -----  
Bounding box area(px): 2100  
Area (px): 54042.0  
Words : 110  
Mean gray-level value in bounding box: 56145.631428571425

-----Region 2 -----  
Bounding box area(px): 34090  
Area (px): 39961.0  
Words : 56  
Mean gray-level value in bounding box: 2433.173217952479

-----Region 3 -----  
Bounding box area(px): 19084  
Area (px): 269889.0  
Words : 397  
Mean gray-level value in bounding box: 29429.640222175643

-----Region 4 -----  
Bounding box area(px): 20655  
Area (px): 74261.0  
Words : 99  
Mean gray-level value in bounding box: 8145.178503994191

-----Region 5 -----  
Bounding box area(px): 112  
Area (px): 6264.0  
Words : 21  
Mean gray-level value in bounding box: 303263.7589285714

-----Region 6 -----  
Bounding box area(px): 1701  
Area (px): 94505.0  
Words : 172  
Mean gray-level value in bounding box: 126993.63374485596

-----Region 7 -----  
Bounding box area(px): 4992  
Area (px): 2778.0  
Words : 9  
Mean gray-level value in bounding box: 2077.21233974359

-----Region 8 -----  
Bounding box area(px): 2475  
Area (px): 435.0  
Words : 1  
Mean gray-level value in bounding box: 488.27272727272725

-----Region 9 -----  
Bounding box area(px): 5292  
Area (px): 34944.0  
Words : 57  
Mean gray-level value in bounding box: 13260.129818594105

-----Region 10 -----  
Bounding box area(px): 646  
Area (px): 5139.0  
Words : 4  
Mean gray-level value in bounding box: 12424.482972136222

-----Region 11 -----  
Bounding box area(px): 1590  
Area (px): 331.0  
Words : 1  
Mean gray-level value in bounding box: 645.5264150943397

## Ενδεικτικός Αλγόριθμος Projection Profile

**Περιγραφή:** Η μέθοδος projection profiles βασίζεται στην ιδέα ότι μπορούμε να ανιχνεύσουμε περιοχές κειμένου σε μία εικόνα υπολογίζοντας το πλήθος των "μαύρων" pixels κατά μήκος είτε της οριζόντιας είτε της κάθετης κατεύθυνσης. Αυτό γίνεται σε μια **δυναμική εικόνα**, όπου τα εικονοστοιχεία του κειμένου έχουν τιμή 1 (μαύρο) και του φόντου τιμή 0 (λευκό).

---

### Οριζόντιο Projection Profile

1. Είσοδος: μία **δυναμική εικόνα**, όπου το φόντο έχει τιμή 0 και τα γράμματα τιμή 1.
  2. Για κάθε **γραμμή** της εικόνας (δηλαδή για κάθε τιμή  $y$  στον κατακόρυφο άξονα):
    - Υπολόγισε το **άθροισμα όλων των pixels** αυτής της γραμμής.
  3. Αν το άθροισμα είναι **μεγάλο** → η γραμμή περιέχει πιθανότατα κείμενο.  
Αν το άθροισμα είναι **μηδέν ή πολύ μικρό** → πιθανώς είναι κενή περιοχή (διάστημα μεταξύ υποπεριοχών).
  4. Εντόπισε διαστήματα ( $y\_start$ ,  $y\_end$ ) όπου το άθροισμα ξεπερνά ένα κατώφλι → αυτά αντιστοιχούν σε **υποπεριοχές με κείμενο**.
  5. Για κάθε τέτοια περιοχή, κατασκεύασε ένα **περιβάλλον κουτί (bounding box)** που περιλαμβάνει όλες τις στήλες (από  $x=0$  έως  $x=πλάτος$  εικόνας).
- 

### Κατακόρυφο Projection Profile

Ίδια διαδικασία αλλά κατά **στήλη**.

1. Είσοδος: μία **δυναμική εικόνα**, όπου το φόντο έχει τιμή 0 και τα γράμματα τιμή 1.
2. Για κάθε στήλη (κάθε τιμή  $x$ ):
  - Υπολόγισε το **άθροισμα των pixels** της στήλης.
  - Αν το άθροισμα είναι μικρό ή μηδέν για πολλές συνεχόμενες στήλες → υποδηλώνει πιθανό **κενό μεταξύ λέξεων**.
3. Μετρώντας τα κενά, μπορείς να εκτιμήσεις **τον αριθμό των λέξεων**