DRAFT -- CAS CS 585 - Spring 2021

The assignment is due 11:59 PM (at midnight) EST, Wednesday, February 17, 2021.
Please scan your written assignment (Part 1) as **a single PDF file and use GradeScope** to submit.
Instructions how to submit the programming assignment will follow.
You are encouraged to work in teams of two or four students. Each student must submit his or her own solutions and acknowledge any teammates.

Part 1: Paper-and-pencil Assignment: 2nd Moments NCC, Evaluating Binary Classifiers of Videos

Please write legibly if you use paper and pencil. You may also type your solutions.

Exercise 1: Using 2nd moments for binary image analysis (Use material from lectures on Feb. 2 and 4)

1. Given four objects with different shapes, please calculate the following values respectively for each object.
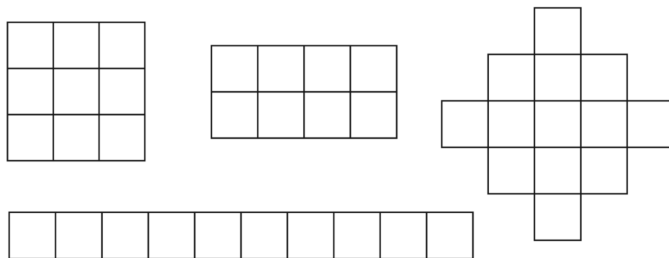
   - Location: Centroid. Explain how you will assign object location when the centroid falls in between pixels.
   - Circularity: E_min/E_max
   - Orientation: Direction of axis of least inertia. Explain how you will assign object direction when this axis is ambiguous.
2. The coordinate origin of an image may be at the lower left corner or at the center of an image. Does the location of the coordinate origin change your answers above? Please answer yes or no and provide a 1-sentence explanation to your answer.
3. The objects shown above are images with just a few pixels. Assume your images have objects of the same shape but much larger size. Are any of the three measures above "invariant to size," which means yield the same value (ignore small round-off errors)? Please answer yes or no and provide a short explanation.

Exercise 2: NCC (wait until lecture on Feb. 9, 2021)

1. Show that the normalized correlation coefficient *r* is invariant to linear brightness changes in the scene *s* (or template *m*). This means you need to prove *r(m,s) = r(m,as+b)* for images *m* and

s, and some constants *a* and *b*.

We defined the normalized correlation coefficient in class as *r= 1/n Σ_i [(s_i - mean(s)) (m_i-mean(m)) / (σ_s σ_m)]*,
where *s_i* and *m_i* are the respective brightness values of the *i*th pixel, mean(m) and σ_m are mean and standard deviation of all pixels in the template and mean(s) and σ_s are mean and standard deviation of all pixels in the sub-image of the scene.

2. Explain why the linear invariance property of the normalized correlation coefficient, shown in part (a), could be useful for image analysis (< 2 sentences).
3. The range of the NCC is between -1 and 1: -1 <= r <= 1. Explain why this property could be useful for image analysis (< 2 sentences).
4. The expected value of the NCC is 0: E[r]=0. Explain why this property could be useful for image analysis (< 2 sentences).

Exercise 3: Evaluating Binary Classifiers of Videos (wait until lecture on Feb. 9, 2021)

Suppose that we have a test dataset of ultrasound videos of the lungs of 200 babies -- 120 have pneumonia and 80 do not have pneumonia. We use it to test the performance of a 2-class video classification model that predicts which baby has pneumonia. Our experiment produces the following confusion matrix.

Confusion Matrix		True Class	
		Pneumonia	Healthy
Hypothesized Class	Pneumonia	64	24
	Healthy	X	Y

1. What is X?
2. What is Y?
3. Compute recall.
4. Compute precision
5. Compute accuracy.
6. Is this classifier specific for pneumonia? Explain.
7. Is this classifier sensitive to pneumonia? Explain.
8. Missing a diagnosis for pneumonia could mean that the baby dies. Providing a false alarm of a healthy baby typically means additional imaging is ordered, e.g., an Xray. Considering this background information, which of the properties of the binary classifier do you propose is most important to improve?

For those of you interested in research: We are currently developing a lung US video classifier for babies in Zambia. We are working with Zambian doctors who are in the process of providing 200 videos for us. U.S. radiologists in our team provide annotations of regions in the lungs that show pneumonia ("consolidations"). In Zambia, an under-resourced country, a clinic is much more likely to have an ultrasound device (the cheapest go for $2,000) than an xray machine (tens of thousands dollars). When there is no image analysis available at a rural clinic (hardware and interpretation expertise), babies typically receive antibiotics. This has led to antibiotics becoming more and more ineffective. AI-supported US video analysis may lead to fewer prescriptions of antibiotics for babies who do not need them while ensuring that sick babies can be diagnosed and given antibiotics immediately.

We meet every Wednesday morning 8:30-9:30 am on zoom if you want to listen. I can provide the link for truly interested students.

Part 2: Programming Assignment (you may start -- some tools will be introduced in the lectures on Feb 9 & 11)

In this assignment, you will build upon the ideas you learned in class and in the labs. Your team will design and implement algorithms that recognize hand shapes or gestures, and create a graphical display that responds to the recognition of the hand shapes or gestures.

Each team must write a webpage report on the assignment. You may use our [template.](#) We provide this template for guidance on content -- you may develop your own html style file. To ensure that each team member builds his or her own electronic portfolio, we ask that everybody submits his or her own report to GradeScope.

Learning Objectives

1. Read and display video frames from a webcam
2. Learn about tracking by template matching
3. Learn about analyzing properties of objects in an image, e.g. object centroid, axis of least inertia, shape (circularity)
4. Create interesting and interactive graphical applications

Requirements

Design and implement algorithms that recognize hand shapes (such as making a fist, thumbs up, thumbs down, pointing with an index finger etc.) or gestures (such as waving with one or both hands, swinging, drawing something in the air etc.) and create a graphical display that responds to the recognition of the hand shapes or gestures. For your system, you are encouraged to try out some of the following computer vision techniques that were discussed in class and use at least a couple of techniques (in particular, binary object shape analysis):

1. horizontal and vertical projections to find bounding boxes of "movement blobs" or "skin-color blobs"
2. size, position, and orientation of object of interest
3. circularity of object of interest
4. template matching (e.g., create templates of a closed hand and an open hand)
5. background differencing: $D(x,y,t) = |I(x,y,t)-I(x,y,0)|$
6. frame-to-frame differencing: $D'(x,y,t) = |I(x,y,t)-I(x,y,t-1)|$
7. motion energy templates (union of binary difference images over a window of time)
8. skin-color detection (e.g., thresholding red and green pixel values)
9. tracking the position and orientation of moving objects

You may use OpenCV library functions in your solution. If you do so, you must understand the OpenCV function in detail -- both the mathematical formulation and the algorithm. In particular, you must be able to explain the function on a whiteboard without access to the OpenCV help pages.

Your algorithm should detect at least four different hand shapes or gestures.

In your report, create a confusion matrix to illustrate how well your system can classify the hand shapes or gestures. You are also asked to create a graphical display that responds to the movements of the recognized gestures. The graphics should be tasteful and appropriate to the gestural movements. Along with the program, submit the following information about your graphics program:

1. An overall description
2. How the graphics respond to different hand shapes and/or gestures
3. Interesting and fun aspects of the graphics display

STILL TO BE FINALIZED: Submission The programming assignment (code), along with the webpage report (html) and results (video), should be submitted to Gradescope.

Margrit Betke, Professor
Computer Science Department
Email: betke@cs.bu.edu