CS365: Artificial Intelligence End Semester Examination

Time: 2 hours 100 marks 2013apr

Question 1. **(Vision)** [5 × 5]

(a) A pinhole camera is moving laterally, as if looking out of a train window. Compute the equations for the optical flow. What do you observe?

(b) *Stereo.* Consider a point object P at depth Z from two pinhole cameras separted by the baseline b. What is the angular disparity ($\delta\theta$) between the two cameras when P moves to $Z + \delta Z$?

(c) A gun manufacturer wishes to locate targets upto 5km away with a stereo system. Targets distance must be accurate to 1m for the gun to be effective. Each camera has a field of view of 45° and a 1000×1000 sensor. What baseline separation would you suggest?

(d) Discuss how i) Harris Corner Detectors, or ii) SIFT are / are not scale independent.

(e) Optical flow, as well as stereo, requires you to solve the correspondence problem. Describe why SIFT (and SIFT like) features are important for correspondence.

Solution:

a. We have image x coordinate $u = -f\frac{X}{Z}$. plane). The camera is looking sideways - i.e. it is moving along X. The camera speed of $\dot{X}$ is the same as the world moving with $-\dot{X}$. Now we have:

$$\frac{du}{dt} = -f\frac{1}{Z}\frac{dX}{dt} = f\frac{\dot{X}}{Z}$$

Hence. the object image will be moving along u with a velocity inversely proportional to Z or the distance from the camera.

b. Here the angle difference between the two images, θ is approximately $= \frac{b}{Z}$, so $\frac{\delta\theta}{\delta Z} = -\frac{b}{Z^2}$, i.e. as P moves to $Z + \delta Z$?, the angle subtended shrinks by $-\frac{b}{Z^2}\delta Z$, which is the angular disparity.

c. The minimum angle discernible in these cameras correspond to 1 pixel = 45°/1000 or .00067 radians. For Z=5000m and δZ = 1m, we have $b = 5000^2 m^2 \times 0.00067 rad/1m$. This gives the unbelievable figure of $b = 16.8km$, which means that it is simply not feasible to use stereo for estimating such large depths, since the effectiveness of stereo drops quadratically with Z.

d. Harris corner detector is rotation invariant and can handle some degree of illumination variation. However, at large scales, it will see a corner as an edge. In SIFT, the scale is first identified by computing the difference of Gaussians (strength of edge) at multiple scales; the features are then computed at the scale for which DoG was maximum. This ensures scale invariance.

e. In optical flow, one has to compute the image motion $\dot{u}$ corresponding to world motion; similarly in stereo one has to compute the δu. Both require that in the new frame, we find the image coordinates that corresponds to the object that was at u in the earlier frame (the differnce between these positions is δu). This is the **correspondence problem**. Since SIFT and HOG are quite robust to scale and orientation, they also work for a range of variation in the views. This enables us to use them for finding correspondences, even where the disparity in views may be quite large.

Question 2. **(Search : Row of tiles puzzle)** [5+6+5]

RRR□WWW represents a row with n red tiles, a blank space, and another n white tiles (here $n = 3$). A tile may move to the blank position if it is next to it, or by jumping over a tile. The goal is to have all the White tiles at the left, and the Red ones at Right.

(a) Model the puzzle as a production sytem with rules describing the legal transitions. Apply each of your transition rules to the state WWRWR□R.

(b) Here are two estimates for the minimum number of moves needed to solve the puzzle from any

state. Discuss if i) each of these results in an admissible heuristic, and ii) which may be expected to result in lesser search :

1. h1: number of tiles not in position
2. h2: sum of distances of tiles from nearest target position.

(c) Construct an admissible heuristic (may or may not be based on of the above). Apply it to the state RRWRW□W for three moves, indicating how your heuristic guides you to your decision.

Solution:

a. The state is represented by a string composed from n Rs, n Ws and a single □. Let X, Y represent any strings, and a,b,c,d be variables $\in R, W, \phi$, where ϕ is a position outside the string. Given that the tile being moved is not blank (in parenthesis for each rule), the transition rules (named after the motion of the blank) be written as:

L1 ($c \neq \phi$): $Xab\square cdY \longrightarrow Xa\square bcdY$	L2 ($d \neq \phi$): $Xab\square cdY \longrightarrow X\square bacdY$
R1 ($b \neq \phi$): $Xab\square cdY \longrightarrow Xabc\square dY$	R2 ($d \neq \phi$): $Xab\square cdY \longrightarrow Xabdc\square Y$

Note: the constraints can also be modeled in other ways - e.g. the blank may not move beyond the string, or by having $a, b \in R, W$, and positing ten additional rules with $b\square cdY$ or $Xab\square$ etc. on the RHS [why 10?].

Applying this to the state WWRWR□R, we have:

L1: WWRWR□R ⟶ WWRW□RR	L2: WWRWR□R ⟶ WWR□RWR
R1: WWRWR□R ⟶ WWRWRR□	R2: cannot be applied since $d = \phi$

b. A heuristic $h(n)$ is admissible if for all nodes n it is not less than the minimal path to the goal $h^*(n)$. In h1, the number of misplaced tiles will require at least one move each to get to the target zone, hence h1 never overestimates h*.

While this is true most of the time for h2 as well, there are a few (very few) positions, such as $W_1W_2\square R_1W_3R_2R_3$, where h2 returns 3 (distance of W_3 is 2, and R_1 is 1) whereas h* is 2. Thus, h2 is not admissible.

At the same time, h2 is almost always admissible, and is considerably "better" informed than h1 - e.g. for the start node RRR□WWW, h1 = 6 whereas h2 = 18, so h2 can guide many finer steps. So I have a strong feeling that although h3 is not admissible it will in practice evaluate many fewer nodes than h1. But I really can't be certain without wider testing - e.g. se the answer to c below, where h2 does not seem to be conferring much of an advantage.

c. Using h1 is our admissible heuristic. For RRWRW□W the cost from start g() = 3, and h1 = 5. The followup states are :

L1: RRWR□WW : h1 = 5 (h2 = 8 + 7 = 15)
L2: RRW□WRW : h1 = 4 (h2 = 7 + 6 = 13)
R1: RRWRWW□ : h1 = 5 (h2 = 8 + 5 = 13)

Here h1 clearly indicates the move L2. [h2 is equivocal between L2 and R1; but R1 would be a disaster since it is a dead-end with no further moves without backtracking.

Step1: L2 to RRW□WRW

In the next move, we start from RRW□WRW, and have the following possibilities:

L1: RR□WWRW : h1 = 6
L2: R□WRWRW : h1 = 3
R1: RRWW□RW : h1 = 4
R2 (goes back to previous state)

Step 2: we choose L2 again, going to R□WRWRW. Now:

L1: □RWRWRW : h1 = 3 (h2: 4+6=10)
R1: RW□RWRW : h1 = 3 (h2: 5+6=11)
R2: (goes back)

Step 3: may choose either L1 or R1. Here h2 prefers R1. However, the next choice from both states may be WR□RWRW after which both routes are equivalent.

So for this example, h2 does not seem to be providing any real benefit.

Question 3 **(Logic):** [4 + 10 + 5+ 5]

(a) given the Propositions:
A : *Today is Wednesday* ;
B : *This is a question about Vision* ;
C : *It snowed in Kanpur yesterday.*

1. Mark A,B,C as True / False (as in this present context).
2. Now, which of the following are true:
   1. $B \Rightarrow C$ 2. $\neg C \Rightarrow A$ 3. $C \Rightarrow (A \wedge B)$ 4. $((C \vee A) \wedge \neg B) \wedge \neg A$

(b) Construct a Resolution refutation proof for the following.

Predicates:			
	HE(x) : *x hates exams*	1. All students hate exams	$\forall x\ (S(x) \Rightarrow HE(x))$
	P(x) : *x is a professor*	2. Some professors are also students	$\exists x\ (P(x) \wedge S(x))$
	S(x) : *x is a student*	$\therefore$ Some professors hate exams	$\exists x\ (P(x) \wedge HE(x))$

(c) Is the statement "All students hate exams" true over all the students you know? Indicate your opinion on these two statements: a) Universal statements are always too strong. b) Boolean Logic is too inflexible for modeling human reasoning.

(d) Express the statement "X% students hate exams" as a conditional probability. Do the same for "Y% of profs are also students". Assign values to X and Y. Now, what can you say about professors hating exams?

Solution:

a. 1. In the context during the exam, A is True, B and C are false. 2. 1. True 2. True 3. True 4. False

b. Clause form:
1. $\neg S(x) \vee HE(x)$
2. $P(A)$
3. $S(A)$
4. $\neg P(x) \vee \neg HE(x)$

here statement 2 is broken into two parts, 2 and 3, with the same skolem constant in each.

Proof by resolution refutation:
5. $\neg HE(A)$ (4,2, s = x/A)
6. $\neg S(A)$ (5,1, s = x/A)
7. NIL (6,3) □

c. There may be a few students - may be as much as 10% - who actually quite like exams - it's a sort of challenge, gives a little spice to life, etc. Well, it could be fewer, but it is probably not zero. this addresses the (a) part, problems with universal quantifiers.

The (b) part of the q. asks if boolean logic itself is too strong. Here we can see that even the students who don't really *like* exams, may not quite *hate* them. There are many in-between areas between *love* and *hate* - *can-live-with*, *dont mind*, *necessary-evil* etc. If we were to make a questionnaire, a better question to ask may be "to what extent do you hate exams" - 1 - I love em, to 10: I detest them from the bottom of my heart.

So quite possibly life is not quite black and white as boolean logic makes it out to be. An universal quantifier on top of that is perhaps too strong for modeling human thinking.

d. If we say that 80% of students hate exams, we get:

$$p(HE(x)|S(x)) = 0.8$$

similarly, perhaps

$$p(S(x)|P(x)) = 0.6$$

says that 60% of profs are also students. [Note that this assigns a directionality to the relation - being a student is "dependent" on x being a Prof. On the other hand, the existential statement $\exists x\ (P(x) \wedge S(x))$ is non-directional. This arises because the initial statement - *Some professors are also students* - does have a directionality to it, which was lost in the existentialization process.

However, you cannot make any inference about professors hating exams from this data. This is most clear if you view the conditional probabilities as Venn diagrams - 80% of the area of S is taken up by HE, though much of it may be outside. The area HE-S = lots of non-students may hate exams - the lecture hall staff, the kids who have to keep quiet, mothers of doting kids, etc. Similarly 60% of the Professor pie is taken up by S. However, S is probably a much bigger pie than P; so we can't really tell how much of the P circle is taken up by HE, i.e. $p(HE(x)|P(x))$.

Question 4. **(Projects)** [7 × 5]

(a) SIFT or HOG descriptors can be viewed as downsampled Distribution Fields. What feature on the DF is focused on while mapping these to HOG or SIFT?

SOLUTION:
A distribution field is a multi-layer representation of the image obtained by quantizing a desired feature (e.g. intensity) into k bins, resuling in a kx image-size representation. For vector features (e.g. gradient) there may be kxk layers.
SIFT and HOG are defined on the gradient space on the image; they can equivalently be thought of as gradients on the DF as well. However, since the DF is bounded, the accuracy of the model may not be very good.

(b) In the Adaboost algorithm, how are the weak classifiers selected, and how are these combined to create a strong classifier?

SOLUTION:
Each weak classifier is initially selected based on their ability to return low False Negative numbers (does not reject when the object is present). In Adaboost, a cascade is formed of these. Other general boosting ideas such as giving higher importance to objects misclassified by earlier classifiers are also adopted.
In the face detection scenario the classifiers are based on Haar-like rectangular features.

(c) What is HSV? Why would it be preferred over RGB for detecting a single object such as a ball? How can depth be estimated from a single image?

SOLUTION:
HSV or "Hue Saturation Value" is an colour representation that is an alternative for RGB. Here the Hue dimension comes around in a full circle (violet - purple - red), which is not true if you think of colour as frequencies. The main advantage of HSV is that it is somewhat independent of intensity. So the colour of the ball will look roughly the same whether it is in shadow or in light. The depth may be inferred by considering the image size. Based on some calibration images where distances are known, distance in a new image can be estimated. ,

(d) What does GDL stand for? Can you suggest a transition rule in GDL for the *Row of tiles* puzzle from Q2?

SOLUTION:
GDL = Game Descriptive Language.
For the Row of tiles, we would need:
a) INIT statements: 7 statements such as (INIT, 1, R) ... [let 1 to 7 be the row positions.

b) NEXT statements: These tell you where the player can play a tile given the current state. These can have three parts: CELL : what must be the board state for the rule to fire - e.g. (CELL ?A ?B BLANK ?C ?D] DOES 0PLAYER (LEFT2) [or some other move] and finally (TRUE (CELL BLANK ?B ?A ?C ?D))) which gives the resultant state.

(e) What is a parse of a natural language expression? Try explaining it for a sentence such "Take a left at the second junction".

SOLUTION:
Given the parts of speech of the words in the sentence, a parse breaks it into PHRASES. It uses a probabilistic grammar (in this case a dependency grammar) for doing this.
e.g. "a left at the second junction" is an NP, with the PP "at the second junction" complementing the NP "a left". the entire sentence here is a Verb phrase (it is an imperative), with the verb "take", followed by the long NP.

(f) Explain the terms i) wordnet, ii) synset, iii) hypernym iv) sense-tagging
Wordnet = enriched dictionary with all the words of a language, organized in terms of similar meaning groupings called synsets.

SOLUTION:
Synset: groupings of words or multi-word expressions having the same meaning, e.g. "enter", "go in", etc. would be a verb synset. Each synset roughly corresponds to a concept in the world.
Hypernym: Synset A is a hypernym of B if B is a type of A. e.g. vehicle is a hypernym of car.
Sense-taggin: assigning the correct synset for a word which may appear in many synsets. e.g. "take" in the sentence above, may mean choose, or it may mean physically take something, etc.

(g) Define TF-IDF. Given two documents, suggest how TF-IDF may be used to identify the correlations between the documents.

SOLUTION:
TF = how many times a word w appears in the document d; may be normalized IDF = number of documents / number of documents in which word w appears. (usually log of this number is taken).
TF-IDF = TF $\times$ IDF
For similarity between two documents, find the TF-IDF vector of every word in each document ; this is the vector corresponding to the document. Then similarity is the cosine distance.