NATIONAL UNIVERSITY OF SINGAPORE

EXAMINATION FOR

(Semester I: 2017/2018)

EE5731 – VISUAL COMPUTING

November / December 2017 - Time Allowed: 2.5 Hours

INSTRUCTIONS TO CANDIDATES:

1. This paper contains **FIVE (5)** questions and comprises **SIX (6)** printed pages.
2. Answer all questions.
3. This is a CLOSED BOOK examination but candidates are allowed to bring in **ONE (1)** A4 sheet of paper, on which they may write any information, into the examination hall.
4. Total marks are 100.

Q1:

A. One of the basic steps in SIFT is to reject edge-like responses using the following equation:

$$\frac{\mathrm{Tr}(\mathbf{H})^2}{\mathrm{Det}(\mathbf{H})} < \frac{(r+1)^2}{r}$$

where matrix **H** is a Hessian matrix, and *r* is a scalar value. Matrix **H** is obtained from a *DoG* image. Concerning this equation, discuss how the Hessian matrix can be used to find edge-like response in a DoG image?

[7 points]

B. SIFT descriptor has a length of 128. Assuming that you generate this descriptor from an image:
   i. Explain what information is carried by these 128 numbers
   ii. Justify how these numbers can be invariant to certain changes in the image. You must mention 3 invariances, and justify each of them.

[8 points]

C. Viola-Jones' face detection uses the Haar-like features and not SIFT, though many other computer vision algorithms do use SIFT. Mention 3 advantages of the Haar-like features over SIFT. You must provide justification for each advantage you mentioned.

[10 points]

Q2:

A. Assuming two images are taken from two different cameras, and the center of the first camera coincides with the origin of the world coordinates, while the second camera is placed with some distance to the first camera. Prove mathematically that:

$$\mathbb{F} = [\mathbb{K}_2\mathbf{t}_2]_\times \mathbb{K}_2\mathbb{R}_2\mathbb{K}_1^{-1},$$

where we know that $\mathbb{F} = [\mathbb{P}_2\mathbf{c}]_\times \mathbb{P}_2\mathbb{P}^+$. To be able to answer this, you must know the definition of the camera matrix, which carries the information of the intrinsic and extrinsic camera parameters.

[10 points]

(Q2 continues on page 3)

B. Given camera matrices, $\mathbf{P}_1$ and $\mathbf{P}_2$, and a set of corresponding pixels, $\mathbf{x}$ and $\mathbf{x}$`, our goal is to estimate the 3D point, $\mathbf{X}$, for every pair of $\mathbf{x}$ and $\mathbf{x}$`. However, $\mathbf{x}$ and $\mathbf{x}$` can be affected by noise or outliers, and this causes problem for the triangulation. To solve this problem, we can employ MLE (Maximum Likelihood Estimation), whose cost function can be described as:

$$C(\mathbf{x}, \mathbf{x}') = d(\mathbf{x}, \hat{\mathbf{x}})^2 + d(\mathbf{x}', \hat{\mathbf{x}}')^2,$$

where function d is the Euclidean distance, and the other notations refer to Figure Q2.B. Regarding this, answer the following questions:

i. What is the constraint we should put on $\hat{\mathbf{x}}$ and $\hat{\mathbf{x}}'$ to make sure that the triangulation is well-defined?

ii. Write the MLE formulation that incorporates the cost function above and the constraint on $\hat{\mathbf{x}}$ and $\hat{\mathbf{x}}'$. Justify why your formulation works.

iii. Explain the basic idea of optimizing the MLE cost function.

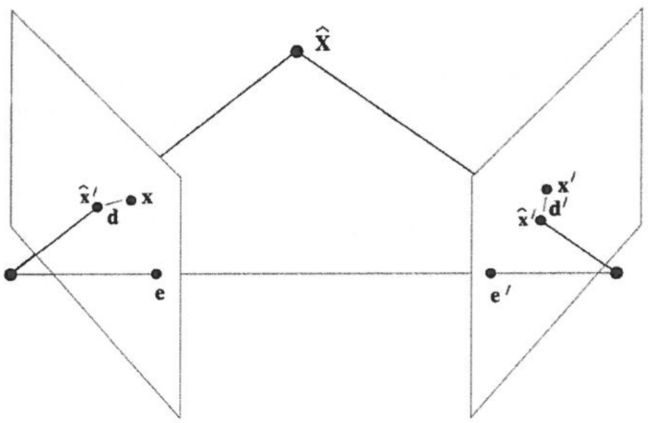

Figure Q2.B. Minimization of geometric error.

[15 points]

Q3:

A. Markov random fields are used in many computer vision algorithms. The objective functions in MRFs normally consist of data terms and prior terms. Explain how the factorization of joint probability distributions affects the definitions of data and prior terms as well as the optimization of an MRF.

[10 points]

(Q3 continues on page 4)

B. Assume you are given a task to detect the pose of a person from a single image by using the pictorial structures. Figure Q3.B shows the example of the output and the tree model used in the recognition. Regarding the model, let's define $L = \{\mathbf{l}_0, \mathbf{l}_1, \ldots, \mathbf{l}_N\}$, where the state of body part i is given by $\mathbf{l}_i = (x_i, y_i, \theta_i, s_i)$; x_i and y_i is the position of the part center in image coordinates, θ_i is the part orientation, and s_i is the part scale (relative to the size of the part in the training set). N indicates the number of the body parts we intend to predict.

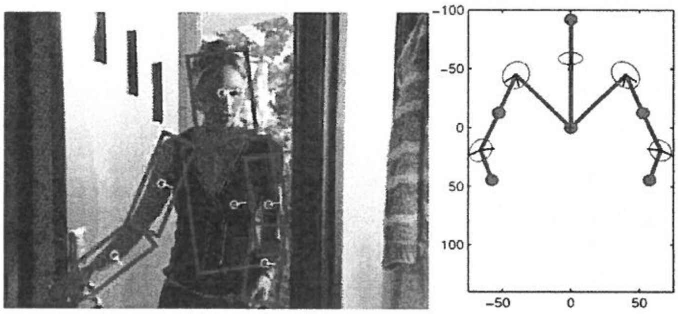

Figure Q3.B. Left: An example of upper body pose recognition. Right: The tree model.

Given the input image, D, and the pose or configuration L,

i. Write the maximum a posteriori (MAP) of the body part prediction using AdaBoost and the tree model.

ii. Justify why your MAP model works. The justification should include how, at the conceptual level, the matching algorithm and spatial relations of the articulated model work.

[10 points]

Note: Below is the AdaBoost algorithm, which might help you answer Q3.B.

- For $t = 1, \ldots, T$:
  1. Normalize the weights, $w_{t,i} \leftarrow \frac{w_{t,i}}{\sum_{j=1}^{n} w_{t,j}}$
  2. Select the best weak classifier with respect to the weighted error
     $$\epsilon_t = \min_{f,p,\theta} \sum_i w_i \, |h(x_i, f, p, \theta) - y_i| .$$
     See Section 3.1 for a discussion of an efficient implementation.
  3. Define $h_t(x) = h(x, f_t, p_t, \theta_t)$ where f_t, p_t, and θ_t are the minimizers of ϵ_t.
  4. Update the weights:
     $$w_{t+1,i} = w_{t,i}\beta_t^{1-e_i}$$
     where $e_i = 0$ if example x_i is classified correctly, $e_i = 1$ otherwise, and $\beta_t = \frac{\epsilon_t}{1-\epsilon_t}$.
- The final strong classifier is:
  $$C(x) = \begin{cases} 1 & \sum_{t=1}^{T} \alpha_t h_t(x) \geq \frac{1}{2} \sum_{t=1}^{T} \alpha_t \\ 0 & \text{otherwise} \end{cases}$$
  where $\alpha_t = \log \frac{1}{\beta_t}$

Q4:

A. According to robust statistics, in optical-flow cost function, L1-norm is more robust than L2-norm. One of the L1-norm functions is Charbonnier, which is expressed as:

$$\rho_d(x) = (x^2 + \epsilon)^a$$

where $a \leq 0.5$. Explain why this function is more robust than an L2-norm function, and provide a simple example to support your argument.

[10 points]

B. Suppose you are given a task to calculate the optical flow from transparent but reflective objects, like the glass shown in Figure Q4.B. Unlike the traditional optical flow, there will be two different flow maps. One represents the motion of the reflected foreground and the other represents the motion of the transmitted background. Note that, the foreground and background motions are independent, and the glass can be assumed static. Moreover, we can also assume an image captured by the camera is a linear combination of the foreground and the background. Concerning this, write cost function that can possibly resolve the problem. You must provide justifications on how and why the cost function works.

[10 points]

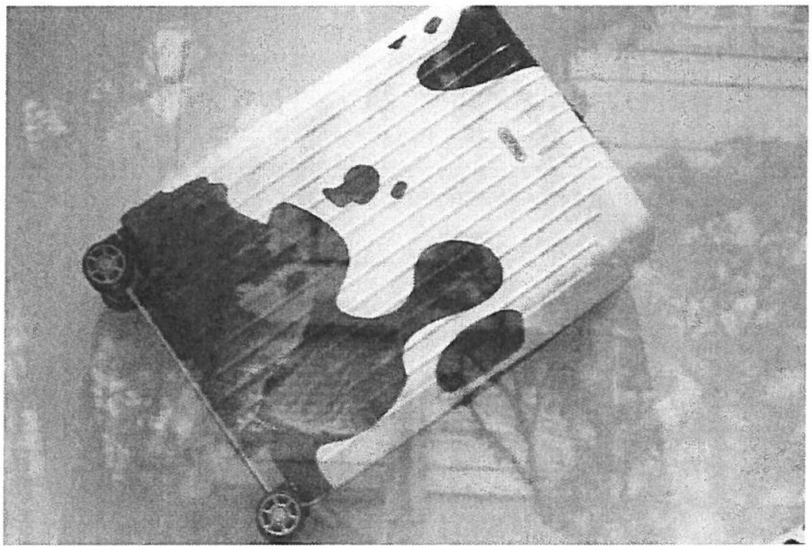

Figure Q4.B. Image of transparent but reflective objects

Q5:

A. One of the methods in image inpainting is the exemplar-based method, where the algorithm is described in Figure Q5. The method distinguishes itself from the other methods by introducing the priority function in step 1b of the algorithm. It consists of the confidence term and data term, with their definitions written respectively as:

$$C(\mathbf{p}) = \frac{\sum_{\mathbf{q}\in\Psi_{\mathbf{p}}\cap(\mathcal{I}-\Omega)} C(\mathbf{q})}{|\Psi_{\mathbf{p}}|}, \qquad D(\mathbf{p}) = \frac{|\nabla I_{\mathbf{p}}^{\perp} \cdot \mathbf{n}_{\mathbf{p}}|}{\alpha}$$

i. What is the intuitive meaning of C(p)?
ii. To obtain D(p), how can we calculate the normal and the isophote?
iii. Why is the priority function able to work efficiently in filling up holes in the input image?

[10 points]

- **Extract the manually selected initial front $\delta\Omega^0$.**
- **Repeat until done:**

1a. **Identify the fill front $\delta\Omega^t$. If $\Omega^t = \emptyset$, exit.**
1b. **Compute priorities $P(\mathbf{p})\quad \forall \mathbf{p} \in \delta\Omega^t$.**
2a. **Find the patch $\Psi_{\hat{\mathbf{p}}}$ with the maximum priority,**
i.e., $\hat{\mathbf{p}} = \arg\max_{\mathbf{p}\in\delta\Omega^t} P(\mathbf{p})$.
2b. **Find the exemplar $\Psi_{\hat{\mathbf{q}}} \in \Phi$ that minimizes $d(\Psi_{\hat{\mathbf{p}}}, \Psi_{\hat{\mathbf{q}}})$.**
2c. **Copy image data from $\Psi_{\hat{\mathbf{q}}}$ to $\Psi_{\hat{\mathbf{p}}}\ \forall \mathbf{p} \in \Psi_{\hat{\mathbf{p}}} \cap \Omega$.**
3. **Update $C(\mathbf{p})\quad \forall \mathbf{p} \in \Psi_{\hat{\mathbf{p}}} \cap \Omega$**

Figure Q5. Inpainting Algorithm

END OF PAPER