Université Clermont Auvergne

Exercice: Master et ecole d'ingénieur

Apprentissage : Reconnaissance de fleurs

La démarche suivie, la justification des résultats obtenus et la présentation interviendront pour part importante dans la note finale.

Introduction

Cet exercice concerne l'étude d'une méthode basée apprentissage pour la reconnaissance de variétés de fleurs. La figure 1 montre des exemples d'images acquises et la figure 2 les trois catégories de fleurs que l'on souhaite reconnaitre. On dispose d'une caméra et d'une unité de traitement qui retourne un ensemble $\mathcal{X} \doteq \{\mathbf{x}_k\}_{k=1,..,K}$ d'objets détectés avec $\mathbf{x}_k \doteq (h,s)^T$ les composantes moyennes de teinte et de saturation de l'objet k.

D'autre part, on dispose d'un ensemble de données annotées par un expert : $\mathcal{X}^A \doteq \{(\mathbf{x}_k^a, y_k^a)\}_{k=1,..,K^a}$ avec $y_k^a \in \{1;2;3\}$, un label de catégorie d'objets tel que $y_k^a = 1$ pour une fleur de la variété 1, $y_k^a = 2$ pour une fleur de la variété 2 et $y_k^a = 3$ pour une fleur de la variété 3.

On souhaite étudier une méthode de classification automatique qui associe un label y à un objet inconnu $\mathbf{x}$.

Figure 1 – Exemples d'images acquises

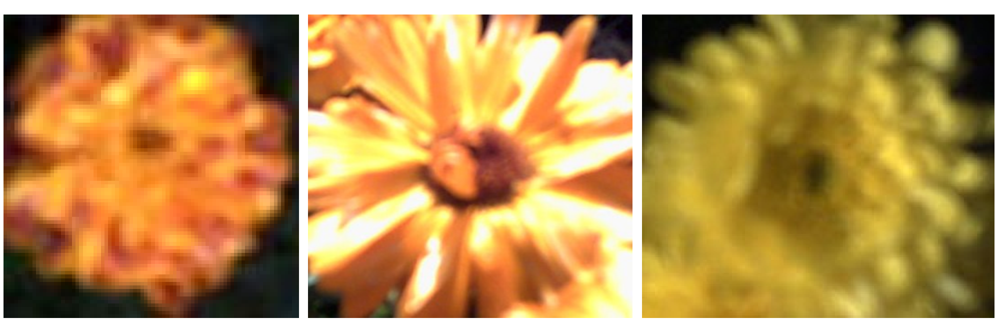

Figure 2 – Exemples de fleurs détectées par le module de détection

1 Classification bayesienne

On propose une formalisation du problème par une technique de classification Bayesienne.

Question 1. Expliquez le principe de la classification Bayesienne et donnez le nom et l'expression de la formule permettant de calculer les distributions à postériori à partir des probabilités à priori et des vraisemblances.

Des statistiques montrent qu'en fonction de la zone géographique (zone1, zone2, zone3), les proportions d'occurrence de chaque type de fleur sont différentes. Par exemple, dans le cas d'un environnement de zone 1, 20% des objets observés sont des fleurs de catégorie 1, 40% sont des fleurs de catégorie 2 et 40% sont des fleurs de catégorie 2. L'information d'environnement est fournie par un GPS. Dans le cas d'une défaillance du GPS, on considérera des proportions équiprobables pour les trois classes (types de fleurs).

Question 2. Comment prendre en compte l'information donnée par le GPS dans la formule de la question 1 ?

Une mesure $\mathbf{x}$ donne les vraisemblances suivantes : $p(\mathbf{x}|y=1)=0,6$, $p(\mathbf{x}|y=2)=0,3$ et $p(\mathbf{x}|y=3)=0,3$.

Question 3. Calculez les probabilités a posteriori $p(y=1|\mathbf{x})$, $p(y=2|\mathbf{x})$, $p(y=3|\mathbf{x})$ dans le cas d'un environnement en zone 1

Question 4. Calculez les probabilités a posteriori $p(y=1|\mathbf{x})$, $p(y=2|\mathbf{x})$, $p(y=3|\mathbf{x})$ dans le cas d'une défaillance du GPS

2 Classification non paramétrique : méthodes de K plus proches voisins

On cherche maintenant à appliquer une classification non paramètrique de type K plus proches voisins

Question 5. Rappeler le principe de cette méthode de classification.

On définit une fonction de mesure (norme L2) entre la taille de deux objets détectés. On a ainsi le tableau suivant, pour deux fleurs de classe inconnue $\mathbf{x}_1$ et $\mathbf{x}_2$: mesures entre l'objets et un ensemble d'objets annotés, issus de $\mathcal{X}^A$:

	$\mathbf{x}_1^a, y_1^a=1$	$\mathbf{x}_1^a, y_1^a=1$	$\mathbf{x}_1^a, y_1^a=1$	$\mathbf{x}_1^a, y_1^a=1$	$\mathbf{x}_1^a, y_1^a=2$	$\mathbf{x}_1^a, y_1^a=2$	$\mathbf{x}_1^a, y_1^a=3$	$\mathbf{x}_1^a, y_1^a=3$	$\mathbf{x}_1^a, y_1^a=3$
$\mathbf{x}_1$	0,21	0,3	0,2	0,1	0,19	0,8	0,41	0,33	0,6
$\mathbf{x}_2$	0,8	0,4	0,5	0,51	0,09	0,11	0,55	0,32	0,17

Question 6. Calculez, en utilisant la méthode des trois plus proches voisins, les probabilités a posteriori $p(y=1|\mathbf{x}_1)$, $p(y=2|\mathbf{x}_1)$, $p(y=3|\mathbf{x}_1)$, $p(y=1|\mathbf{x}_2)$, $p(y=2|\mathbf{x}_2)$ et $p(y=3|\mathbf{x}_2)$