



---

## **Guide pédagogique**

***Module « Perception 3D et Interprétation »,  
Option IASD, Spécialité "Image" – 10.2 (3 crédits ECTS)***

---

### ***Place du module et enjeux***

Une grande partie de l'intelligence humaine repose sur notre capacité à voir et analyser ce qui est vu. La perception que nous pouvons avoir de la profondeur de champs, l'appréciation des volumes et l'interprétation des images sont autant de composantes primordiales de cette analyse. De nombreuses applications utilisent des données 3D. Que ce soit des données obtenues par stéréovision ou acquises directement, elles doivent la plupart du temps être fusionnées afin d'obtenir une description globale de la scène considérée. Ce module s'intéresse particulièrement à la 3D et à la reconnaissance d'objets ou l'interprétation de scènes. De l'acquisition à la stéréovision en passant par la perception et l'interprétation, différents outils mathématiques et logiciels seront décrits. La détection d'objets ou la description de scène seront des applications privilégiées.

---

## **Teaching guide and syllabus**

***« 3D Perception and Interpretation » module,  
DSAI option, specialization "Image" – 10.2 (3 ECTS credits)***

---

### ***Subject matter importance and associated issues***

Human intelligence is mainly based on our ability to see and analyze what is seen. The perception we can have of the depth of field, the appreciation of volumes and the interpretation of images are all essential components of this analysis. Whether obtained by stereovision or directly acquired, 3D data must be merged in order to obtain a global description of the scene. This module is particularly interested in 3D and object recognition or scene interpretation. From acquisition to stereovision, perception and interpretation, various mathematical tools and software will be described. Object detection or scene description will be preferred applications.

**Responsable : Philippe MONTESINOS**

Téléphone : 04 34 24 62 69

Courriel : [philippe.montesinos@mines-ales.fr](mailto:philippe.montesinos@mines-ales.fr)



**IMT Mines Alès**  
École Mines-Télécom

<b>ENSEIGNEMENTS ACADÉMIQUES</b>	<b>Volume horaire</b>	Détail des coefficients	Crédits
<b>Perception 3D et Interprétation</b> o 3D, mouvement, interprétation de scènes, reconnaissance d'objets	<b>50 h</b> 50	1	3

**Matière 1**

<b>Titre de la matière : Perception 3D et Interprétation</b>	
<b>Code : 2IA-iasd-img 10.2</b>	<b>Titre du module : « Perception 3D et Interprétation »</b>
<b>Semestre : S10</b>	<b>Cursus de rattachement : département 2IA, options IASD, spécialité Image</b>

Heures présentiel	Heures total	Cours	TD	TP	Projet	Contrôles	Travail personnel	Coef /module	ECTS
50	75	27			21	2	25	1	3

<b>Résumé</b>	<p>Ce module se focalise sur l'exploitation des données obtenues par des méthodes présentées dans le module « Processus Visuels ». Il présente et détaille deux aspects importants de la Vision par Ordinateur : le 3D et la reconnaissance d'objets ou l'interprétation de scènes. Dans un premier temps, il aborde l'acquisition 3D (actif, passif), les notions géométriques pour la stéréovision et le mouvement (géométrie épipolaire), la perception 3D (calibration, reconstruction 3D, méthodes actives et passives), la modélisation d'environnement 3D (SLAM), les notions de réalité virtuelle ou augmentée seront abordées. Dans un deuxième temps les méthodes de reconnaissance d'objets et d'interprétation de scènes feront l'objet d'une attention toute particulière (description de scènes, mise en correspondance symbolique, relaxation, méthodes locales, méthodes globales). Parmi tous les points abordés, nous présenterons notamment les méthodes de mise en correspondance et de détection d'objets par invariants locaux, puis les méthodes basées sur des description globales de scènes (<i>Deep Learning</i>).</p>
---------------	---

<b>Responsable</b>	Philippe Montesinos – LGI2P/IMT Mines Alès
<b>Équipe enseignante</b>	Baptiste Magnier – LGI2P/IMT Mines Alès Philippe Montesinos – LGI2P/IMT Mines Alès Pierre Slangen – LGI2P/IMT Mines Alès

<b>Mots-clés</b>	Mise en correspondance, Reconnaissance d'objets, Deep Learning, Calibration, Reconstruction 3D, Stéréo, Mouvement, Réalité virtuelle, Réalité augmentée.
<b>Prérequis</b>	Linux, langage C/C++, Méthodes d'apprentissage (Deep learning).

<p><b>Contexte et objectif général :</b></p> <p>De nombreuses applications utilisent des données 3D. Que ce soit des données obtenues par stéréovision ou acquises directement (il existe des caméras délivrant directement des données 3D denses), ces données 3D doivent la plupart du temps être fusionnées afin d'obtenir une description globale de la scène considérée. Les méthodes de calibration et de reconstruction 3D sont alors essentielles à la réalisation de ces objectifs (calibration stéréo et reconstruction 3D, mouvement, calibration mixtes stéréo et acquisition 3D actives, Simultaneous Localization And Modelisation : SLAM).</p> <p>Un deuxième domaine d'application important de la vision est lié à la détection et la reconnaissance d'objets ou de scènes. Là encore de nombreuses applications utilisent les techniques de vision par exemple pour localiser un robot dans un environnement (navettes urbaines autonomes, drones, véhicules autonomes robot mobile d'assistance à la personne, ...), ou encore pour détecter et reconnaître automatiquement un produit dans un rayon de supermarché (réalité augmentée et information consommateur sur un produit donné, caisses automatisées, ...). Ce cours à la fois théorique et pratique détaille donc les applications actuelles et fournit aux étudiants les bases pour construire les applications futures.</p> <p><b>Programme et contenu :</b></p> <ul style="list-style-type: none"> <li>- Acquisition 3D (3h de cours)</li> <li>- Calibration, reconstruction 3D, Mouvement, SLAM (12h de cours)</li> <li>- Description et Interprétation de scènes (3h de cours)</li> <li>- Détection et reconnaissance d'objets par invariants locaux (3h de cours)</li> <li>- Reconnaissance d'objets ou de scènes par des méthodes globales (Apprentissage et Deep Learning) (3h de cours)</li> <li>- Applications : Réalité virtuelle / augmentée, Vision industrielle, Robotique, Véhicules autonomes, Drones, Applications Médicales... (3 h de cours 21h de Projet).</li> </ul>
--

**Méthode et organisation pédagogique :**

Les enseignements sont prévus pour 30 élèves. Les exercices pratiques seront réalisés sur des ordinateurs de l'école préconfigurés sous Linux (Langages C/C++/CUDA), Bibliothèques OpenCV, VL2, TensorFlow, YOLO, ...

Le découpage est prévu comme suit :

- 27h de cours
- 21h de TP / Projet
- 2h de contrôle

**Acquis d'apprentissage visés :**

Mettre en œuvre une ou plusieurs méthodes étudiées au travers d'une application.

**Évaluation :** Contrôle écrit (2h, coefficient 1) et Projet (coefficient 1)

**Retour sur l'évaluation fait à l'élève :**

Retour sur le contrôle 3 semaines après la dernière séance

Retour le projet lors de la soutenance de projet.

**Support pédagogique et références :**

1 Polycopié, supports de cours.

## Méthode et organisation pédagogique

Il s'agit d'un enseignement relativement classique avec une partie réalisée en cours magistral et une partie appliquée au travers de TP.

## Modalité d'évaluation

Le niveau d'acquisition des compétences sera évalué selon les exigences suivantes :

N° indicateur	Indicateur
1	connaître les savoirs formels et pratiques du socle des fondamentaux
2	Exploiter les savoirs théoriques et pratiques
3	Analyser, interpréter, modéliser, émettre des hypothèses, et résoudre

## Répartition

Matière	Contrôle	Coefficients	Type de notation	Indicateurs évalués	Chapitres
Perception 3D, reconnaissance d'objets	Devoir sur table	1	Individuelle	Tous	Tous
Perception 3D, reconnaissance d'objets	Projet	1	Binôme	Tous	Tous

## Engagement de l'étudiant, éthique et professionnalisme

La démarche éthique est définie dans le règlement intérieur de l'établissement. Chaque étudiant s'engage à en prendre connaissance et à la respecter.

**Obligation des cours** : Présence obligatoire pour tous à chaque séance

**Nombre d'heures estimées de travail personnel** : pour acquérir les compétences demandées, il est nécessaire que l'étudiant consacre au minimum 1h30 min de travail personnel de compréhension et d'approfondissement par séance de cours.

25h de travail personnel pour approfondir les notions étudiées, définir le périmètre des projets et effectuer les développements associés.

**Nombre d'heures estimées de préparation aux travaux dirigés (TP)** : 1h minimum.

### Pénalité pour retard

Tout travail remis en retard sans motif valable peut être pénalisé de 1 point par jour de retard.

## Équipe enseignante

Nom	Domaine d'expertise	Courriel/Téléphone
Baptiste MAGNIER	Segmentation d'images numériques, détection de contours, diffusion anisotropique, restauration	<a href="mailto:baptiste.magnier@mines-ales.fr">baptiste.magnier@mines-ales.fr</a> 04 34 24 62 90
Philippe MONTESINOS	Vision par Ordinateur : Stéréo-3D, mouvement, Interprétation de scènes, Reconnaissance d'objets, IA.	philippe.montesinos@mines-ales.fr 04 34 24 62 69
Pierre SLANGEN	Laser optique, analyse rapide, optique appliquée, holographie	<a href="mailto:pierre.slangen@mines-ales.fr">pierre.slangen@mines-ales.fr</a> 04 66 78 56 28

<b>ACADEMIC TEACHING</b>	<b>Teaching hours</b>	Coefficients	Credits
<b>3D Perception and Interpretation</b> o 3D, movement, scene interpretation, object recognition	<b>50 h</b> 50	1	3

**Class 1**

<b>Class title: 3D Perception and Interpretation</b>	
<b>Code : 2IA-iasd-img 10.2</b>	<b>Module title : « 3D Perception and Interpretation »</b>
<b>Semester: S10</b>	<b>Classification : CSAI department, DSAI option, Image specialization</b>

Hours of presence	Total hours	Lectures	Work shop	Labs	Project	Testing	Personal work	Coef /module	ECTS
50	75	27			21	2	25	1	3

<b>Summary</b>	This module focuses on the use of data obtained by methods presented in the "Visual Processes" module. It presents and details two important aspects of Computer Vision: 3D and object recognition or scene interpretation. First, it addresses 3D acquisition (active, passive), geometric notions for stereovision and motion (epipolar geometry), 3D perception (calibration, 3D reconstruction, active and passive methods), 3D environment modelling (SLAM), virtual or augmented reality notions will be addressed. In a second step, special attention will be paid to methods of object recognition and scene interpretation (scene description, symbolic mapping, relaxation, local methods, global methods). Among all the points discussed, we will present methods for mapping and detecting objects by local invariants, then methods based on global scene descriptions (Deep Learning).
----------------	--

<b>Head</b>	Philippe Montesinos – LGI2P/IMT Mines Alès
<b>Teaching team</b>	Baptiste Magnier – LGI2P/IMT Mines Alès Philippe Montesinos – LGI2P/IMT Mines Alès Pierre Slangen – LGI2P/IMT Mines Alès

<b>Key words</b>	Mapping, Object Recognition, Deep Learning Calibration, 3D Reconstruction, Stereo, Motion, Virtual Reality, Augmented Reality
<b>Prerequisites</b>	Linux, C/C++ languages, Machine Learning (Deep Learning)

<b>Context and general objective:</b> Many applications use 3D data. Whether obtained by stereovision or acquired directly (there are cameras that directly deliver dense 3D data), these 3D data must be merged in order to obtain a global description of the scene. 3D calibration and reconstruction methods are essential to achieve these objectives (stereo calibration and 3D reconstruction, motion, mixed stereo calibration and active 3D acquisition, Simultaneous Localization And Modelisation: SLAM). A second important field of vision application is related to the detection and recognition of objects or scenes. Here again, many applications use vision techniques for example to locate a robot in an environment (autonomous urban shuttles, drones, autonomous mobile personal assistance robot vehicles, etc.), or to automatically detect and recognize a product in a supermarket radius (augmented reality and consumer information on a given product, automated cash registers, etc.). This course mixing both theoretical and practical aspects, therefore details current applications and provides students with the basis for building future applications.
<b>Programme and contents:</b> <ul style="list-style-type: none"> <li>- 3D acquisition (3 hours of lecture)</li> <li>- Calibration, 3D reconstruction, Movement, SLAM (12h of lecture)</li> <li>- Scene description and interpretation (3 hours of lecture)</li> <li>- Detection and recognition of objects by local invariants (3 hours of lecture)</li> <li>- Recognition of objects or scenes by global methods (Learning and Deep Learning) (3 hours of lecture)</li> <li>- Applications: Virtual / augmented reality, Machine vision, Robotics, Autonomous vehicles, UAVs, Medical applications... (3h lecture and 21 hour of Project).</li> </ul>
<b>Method and pedagogic organisation:</b> The practical sessions will be carried out on the computers of the school (Linux OS, C/C++ language, Cuda, SSE, AVX) with provided libraries (OpenCV, VL2...). The syllabus of the module will be the following: <ul style="list-style-type: none"> <li>- 27 hours of plenary lectures, divided in 2 main chapters.</li> <li>- 21 hours of project</li> <li>- 2 hour for written exam</li> </ul>
<b>Targeted skills or knowledge :</b> Implement one or more methods studied through an application.
<b>Evaluation:</b> Written exam (2h, coef. 1) and Project (coef 1)
<b>Feedback made to the student:</b> Annotated exam copies within three weeks. Feedback during the project defense.
<b>Teaching material and references :</b> 1 Photocopied material – references

## Method and teaching organisation

This will be a classical course with lectures and applications on TD and Project.

## Testing procedures

The student's level of knowledge acquisition will be evaluated according to the following points :

N° Indicator	Indicator
1	To know the formal and practical knowledge constituting the foundation of a given field
2	Exploit theoretical and practical knowledge
3	Analyse, interpret, model, hypothesize and solve problems

### Grading scheme:

Class	Exam	Coefficients	Administration mode	Evaluated Indicators	Chapters
3D Perception and Interpretation	Exam	1	Individual	all	all
3D Perception and Interpretation	Project	1	Group	all	all

## Student commitments, ethics and professionalism

Expectations concerning ethics are defined in the establishment's code of conduct. Each student is expected to know and respect the code of conduct.

**Obligatory presence in classes:** Presence is required all times in lectures and practical sessions.

### Estimated hours of personal study

in order to acquire the required learning level, the student is expected to spend a minimum of 1.5 hour of personal study time per hour spent in class.

25h to realize the project.

**Estimated hours of preparation required for labs/Work Shop:** 1h

### Late penalties

All late assignments without any valid cause may be penalized by the withdrawal of up to 1 point per day after the due date.

## Teaching team

Name	Field of expertise	Email/Phone
Baptiste MAGNIER	Digital image segmentation, edge detection, anisotropic diffusion, restoration.	<a href="mailto:baptiste.magnier@mines-ales.fr">baptiste.magnier@mines-ales.fr</a> 04 34 24 62 90
Philippe MONTESINOS	Computer Vision: Stereo 3D, motion, scene interpretation, object recognition, AI.	<a href="mailto:philippe.montesinos@mines-ales.fr">philippe.montesinos@mines-ales.fr</a> 04 34 24 62 69
Pierre SLANGEN	Optical laser, rapid analysis, applied optics, holography	<a href="mailto:pierre.slangen@mines-ales.fr">pierre.slangen@mines-ales.fr</a> 04 66 78 56 28

## Approbation

Ce guide pédagogique entre en vigueur à compter du 7 janvier 2019.

Il est porté à la connaissance des élèves par une publication sur le site Web de l'école.

<b>Rédaction</b>	<b>Vérification</b>	<b>Validation</b>
L'enseignant responsable du module :  Philippe MONTESINOS	Le responsable d'UE / de département :  Sylvie RANWEZ	Le directeur de l'école, Pour le directeur et par délégation, Le directeur de la DFA / de la DE :  Michel FERLUT