Probabilistic Graphical Models in Computer Vision (IN2329)

Csaba Domokos

Summer Semester 2017

Announcement: Computer Vision Group *

We currently work on various `research topics`:


3D reconstruction


Optical flow

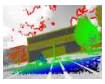

Shape analysis

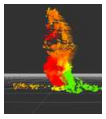

Robot vision


RGB-D vision


Image segmentation


Convex relaxation


Visual SLAM


Scene flow


Deep learning

Inquiries for Bachelor and Master projects are always welcome!

Please complete the form: `https://vision.in.tum.de/application`

1. Introduction

Agenda for today's lecture *

1. Administrative
2. Overview of the course
3. Introduction to Probability Theory
   - Basic definitions
   - Conditional probability, Bayes' rule
   - Independence, conditional independence

Administrative

The course: IN2329 *

The course **Probabilistic Graphical Models in Computer Vision** will be organized as follows:

- **Lectures** will take place between 10:00–12:00 on Mondays in Room 02.09.023.
- **Tutorials** will take place between 14:00–16:00 on Mondays in Room 02.05.014.

Contact details of the lecturer:

Name: Dr. Csaba Domokos
Email: `csaba.domokos@in.tum.de`
Office: 02.09.060

Feel free to contact me!

Tutorials

Participation at the tutorial **is not mandatory, but highly recommended**.

The tutorials combines *theoretical* and *programming* assignments:

- Theoretical assignments will help to understand the topics of the lecture.
- Programming assignments will help to apply the theory to practical problems.

- **Assignment distribution**:
  between 11.00–11.15 on Mondays in Room 02.09.023.
- **Theoretical assignment due**:
  between 11.00–11.15 on Mondays in Room 02.09.023.
- **Assignment presentation**:
  between 14.00–16.00 on Mondays in Room 02.05.014.

To promote team work, please form **groups of two or three students** in order to solve and submit the assignments.

Active students who solve *60% of the assignments* earn a **bonus**.

Exam *

- The exam will be oral (max. **30 mins.**).
- The exam will take place **7th–9th of August**.
- Students **need to be registered** prior to the exam:
  May, 29th–June, 30th via TUM online.

To achieve the **bonus**, the following requirements have to be fulfilled:
Theory:

- **60%** of all theoretical assignments have to be solved.
- The theoretical exercises have to *be presented in front of the class*.

Programming:

- **60%** of all programming assignments have to be presented during the tutorial.

If someone receives a mark between 1.3 and 4.0 in the *final exam*, the mark will be **improved by 0.3 and 0.4, respectively**. (Note that marks of 1.0 and 5.0 cannot be improved.)

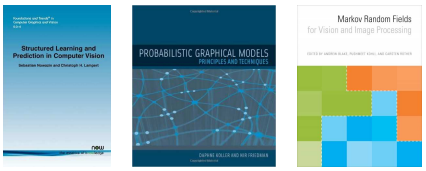
Prerequisites *

The course is intended for *Master students*.

Prerequisites:

- **Basic Mathematics**:
  - *multivariate analysis* and
  - *linear algebra*.
- **Basic Computer Science**:
  - *programming skills* (`Matlab`, `C/C++`) and
  - basic knowledge in *algorithms and data structures* (e.g., dynamic programming).

Recommended literature *



- S. Nowozin, C. H. Lampert. Structured Learning and Prediction in Computer Vision, Foundations and Trends in Computer Graphics and Vision, 2011.
- D. Koller, N. Friedman. Probabilistic Graphical Models: Principles and Techniques, MIT Press, 2009.
- A. Blake, P. Kohli, C. Rother. Markov Random Fields for Vision and Image Processing, MIT Press, 2011.

In addition, we will also mention recent *conference* and *journal papers*.

Course Page *

The `course` page includes general information about the course, where you may also find course materials:

- **Slides for each lecture** (available prior to the lecture)
- **Assignment sheets** (available after the lecture)

The `internal site` of the course page includes

- **solution sheets** (available after the tutorial),
- extra announcements.

Password: *announced on the first lecture*

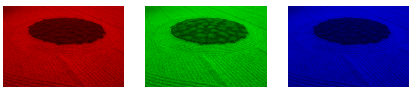
Overview

Overview of the course *

Digital image

A **color digital image** with red, green and blue (RGB) channels can be defined as a function

$$I : \mathcal{D} \subset \mathbb{Z}^2 \to [0, 255]^{3\times N} \subset \mathbb{Z}^{3\times N} \ , \ N = |\mathcal{D}| \ .$$



Red Green Blue

For an image with size of $W \times H$, the domain

$$\mathcal{D} := \{1, \dots, W\} \times \{1, \dots, H\} \ .$$

Zoom in

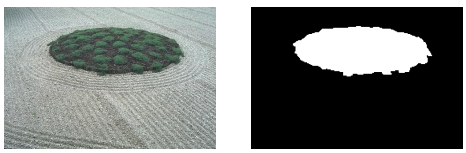
An image I is composed of **pixels** (*picture element* or *picture cell*)

$$i \in \mathcal{D} \ , \qquad I(i) = (r_i, g_i, b_i) \in [0, 255]^3 \ .$$

For more details you may refer to the course on **Computer Vision I: Variational Methods**

Binary image segmentation *

The goal is to give a binary label $y_i \in \mathbb{B} \triangleq \{0, 1\}$ for each pixel i, where 0 and 1 mean the *background* (a.k.a. ground) and the *foreground* (a.k.a. figure), respectively.



Input image Figure–ground segmentation

Source: Berkeley Segmentation Dataset

This task can be formulated as finding a **labeling**:

$$L : \Omega \subset \mathbb{Z}^2 \to \mathbb{B}^N \ , \qquad \mathbf{y} = (y_1, \dots, y_N) \in \mathbb{B}^N \ ,$$

where $|\Omega| = N$ is the number of pixels on the image.

Object detection *

We address the problem of *binary image segmentation*, where we also assume that we are provided with non-local **parameters** that are **known a priori**.

For example, one can assume prior knowledge about the **shape** of the foreground.

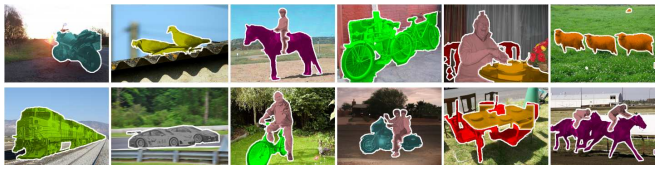
Exemplar binary segmentation of cars assuming shape prior

Source: Lempitsky et al.: Image Segmentation by Branch-and-Mincut. ECCV, 2008.

Semantic image segmentation *

The goal is to give a label $y_i \in \mathcal{L} = \{1, 2, \ldots, c\}$ for each pixel i according to its *semantic meanings*. The *labeling* is defined as

$$L : \mathcal{D} \to \mathcal{L}^N \ .$$

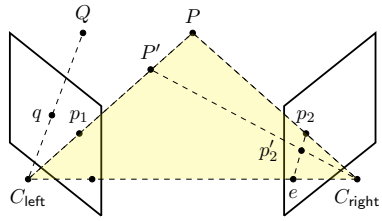


Exemplar semantic segmentations

Source: Xia et al.: Semantic Segmentation without Annotating Segments. ICCV, 2013.

Online demo: `http://www.robots.ox.ac.uk/~szheng/crfasrnndemo`

Stereo matching *

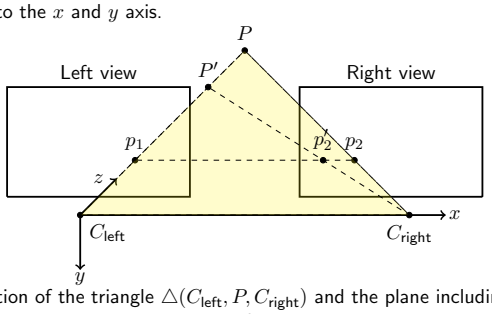


Left view Right view

Given two images (i.e. left and right), an observed 2D point p_1 on the *left image* corresponds to a 3D point P that is situated on a line in $\mathbb{R}^3$. This line will be observed as a line on the *right image*.
P can be determined based on p_1 and p_2. We assume that the pixels p_1 and p_2, corresponding to P, have similar visual appearance.

Rectified cameras *

Suppose that we are given two cameras looking at *parallel direction*. Let C_{left} be the origin of the coordinate system and assume that the *image planes are co-planar* and parallel to the x and y axis.



The intersection of the triangle $\triangle(C_{\text{left}}, P, C_{\text{right}})$ and the plane including the images planes is the segment $\overline{p_1 p_2}$. Therefore $\overline{p_1 p_2}$ is parallel to the x-axis.

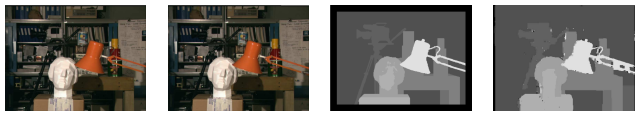
For more details you may refer to the course on **Computer Vision II: Multiple View Geometry**.

Stereo matching *

The goal is *to reconstruct 3D points* according to corresponding pixels. Usually we assume **rectified images** (i.e. the directions of the cameras are parallel), which means that the corresponding pixels are situated in horizontal lines.

Remark that a 3D point P is determined by p_1 and p_2. Therefore, for a given p_1 we want to find the displacement $d_1 \in \mathcal{L} = \{-d, \ldots, 0, \ldots, d\} \subset \mathbb{Z}$ aligning p_1 to p_2. The *labeling* is defined as

$$L : \mathcal{D} \to \mathbb{Z}^N \ .$$

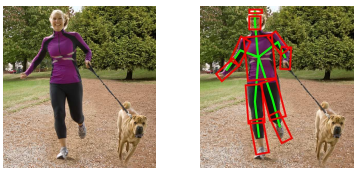


Left view Right view Depth map Result

Source: Komodakis et al.: Approximate Labeling via Graph-Cuts Based on Linear Programming. TPAMI, 2007.

Human–pose estimation *

The goal is to recognize an *articulated object* (e.g., human body) with different connecting parts (e.g., head, torso, left arm, right arm, left leg, right leg).



Input image Human pose estimation

Source: Nowozin and Lampert. Structured Learning and Prediction. 2011.

An object is composed of a number of *rigid parts*, where each part is modeled as a **rectangle**.

Human–pose estimation *

Each part is modeled as a rectangle parameterized by (x, y, s, θ), where

- (x, y) means the **center of the rectangle**,
- $s \in [0, 1]$ is a **scaling factor**, and
- **the orientation** is given by θ.

In overall, we have a four-dimensional *pose space* as *label space* $\mathcal{L}$.

We denote the **label** (i.e. locations) of a part by $l_i = (x_i, y_i, s_i, \theta_i) \in \mathcal{L}$.

An object (e.g., human body) is given by a configuration

$$\mathbf{l} = (l_1, \ldots, l_N) \ ,$$

where l_i specifies the location of **part** v_i and N is the number of object parts.

Probability theory

Reasoning under uncertainty *

We often want to understand a system when we have *imperfect* or *incomplete* information due to, for example, noisy measurement.

There are *two main reasons* why we might **reason under uncertainty**:

- *Laziness*: modeling every detail of a complex system is *costly*.
- *Ignorance*: *lack of complete understanding* of the system.

Probability $P(A)$ refers to a degree of confidence that an event A with uncertain nature will occur.

It is common to assume that $0 \leqslant P(A) \leqslant 1$:

- If $P(A) = 1$, we are certain that A occurs,
- while $P(A) = 0$ asserts that A will not occur.

Experiment, sample space, event

An **experiment** is a (random) process that can be infinitely many times repeated and has a well-defined set of possible **outcomes**. In case of repeated experiments the individual repetitions are also called **trials**.

Example: throwing two "fair dice" (i.e. we assume equally likely chance of landing on any face) with six faces.

The **sample space**, denoted by Ω, is the set of possible outcomes.

Example: $\Omega = \{(i,j) : 1 \leqslant i,j \leqslant 6\}$.

A set of outcomes $A \subseteq \Omega$ is called an **event**. An **atomic event** is an event that contains a single outcome $\omega \in \Omega$.

Example: $A = \{(i,j) : i + j = 11\}$, i.e. the sum of the numbers showing on the top is equal to eleven.

Example *

Suppose that we are capturing a color (RGB) image. Consider each pixel on the image separately as **trials**.

An **outcome** consists of the intensities of a given pixel.

Sample space is given as

$$\Omega = \{(r,g,b) \in \mathbb{Z}^3 \mid 0 \leqslant r,g,b \leqslant 255\} .$$

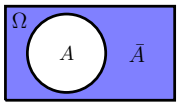
An exemplar **event** A is defined as

$$A = \{(r,b,g) \mid r,g,b \geqslant 128\} .$$

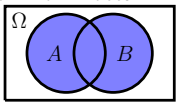
Basic notations

Let A and B be two *events* from an *sample space* Ω. We will use the following notations:

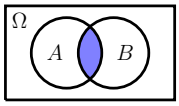
A does not occur: $\bar{A} = \Omega \backslash A$



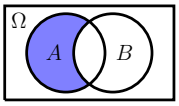
either A or B occur: $A \cup B$



both A and B occur: $A \cap B$



A occurs and B does not: $A \backslash B$



- The $\varnothing$ is called the **impossible event**; and
- Ω is the **sure event**.

Discrete probability space

A *probability space* represents our **uncertainty** regarding an *experiment*.

A triple $(\Omega, \mathcal{A}, P)$ is called a **discrete probability space**, if

- Ω is not empty and **countable** (i.e. $\exists \mathcal{S} \subseteq \mathbb{N}$ such that $|\Omega| = |\mathcal{S}|$),
- $\mathcal{A}$ is the **power set** $\mathcal{P}(\Omega)$ (i.e. the set of all subsets of Ω), and
- $P : \mathcal{A} \to \mathbb{R}$ is a function, called a **probability measure**, with the following properties:
  1. $P(A) \geqslant 0$ for all $A \in \mathcal{A}$
  2. $P(\Omega) = 1$
  3. **σ-additivity** holds: if $A_n \in \mathcal{A}, n = 1, 2, \ldots$ and $A_i \cap A_j = \varnothing$ for $i \neq j$, then
  $$P(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n) .$$

The conditions 1-3. are called **Kolmogorov's axioms**.

Example: throwing two "fair dice" *

For this case a *discrete probability space* $(\Omega, \mathcal{P}(\Omega), P)$ is given by

- **Sample space**: $\Omega = \{(i,j) : 1 \leqslant i,j \leqslant 6\}$, $|\Omega| = 36$.
- $\mathcal{P}(\Omega) = \{\{(1,1)\}, \ldots, \{(1,1),(1,2)\}, \ldots, \{(1,1),(1,2),(1,3)\}, \ldots\}$.
- The (uniform) **probability measure** for all $A \in \mathcal{P}(\Omega)$
$$P(A) = \frac{|A|}{|\Omega|} = \frac{k}{36} ,$$
where k is the number of *atomic events* in A. Such probability space (i.e. all outcomes are equally likely) is called **classical probability space**.

Example: Let A denote the *event* that *"the sum of the numbers showing on the top is equal to eleven"*, that is
$$A = \{(i,j) : i + j = 11\} = \{(5,6),(6,5)\} .$$

Hence
$$P(A) = P(\{(5,6),(6,5)\}) = \frac{2}{36} .$$

Some simple consequences of the axioms

The following rules are frequently used in applications:

- $P(A) = 1 - P(\bar{A}) = 1 - P(\Omega \backslash A)$.

  Proof. Note that A and $\Omega \backslash A$ are disjoint.
  $1 = P(\Omega) = P(A \cup (\Omega \backslash A)) = P(A) + P(\Omega \backslash A)$. □

- $P(\varnothing) = 0$.

  Proof. $P(\varnothing) = 1 - P(\Omega \backslash \varnothing) = 1 - P(\Omega) = 1 - 1 = 0$. □

- If $A \subseteq B$, then $P(A) \leqslant P(B)$.
- $P(A \backslash B) = P(A) - P(A \cap B)$.
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
- $P(A \cup B) \leqslant P(A) + P(B)$.
- ...

(See the proofs in the Exercise.)

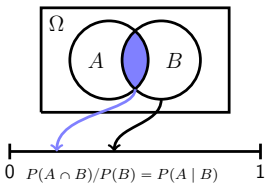
Conditional Probability

Conditional probability

Conditional probability allows us to reason with *partial information*.
If $P(B) > 0$, the **conditional probability of A given B** is defined as

$$P(A \mid B) \triangleq \frac{P(A \cap B)}{P(B)} .$$

This is the probability that A occurs, given we have observed B, i.e. we know the experiment's actual outcome will be in B.



Note that the *axioms and rules of probability theory are fulfilled* for the conditional probability. (e.g., $P(A \mid B) = 1 - P(\bar{A} \mid B)$).

Example *

Consider the problem of *binary segmentation*, that is each pixel belongs to either the foreground or the background.

Let us define a pixel to be "bright", if all its (RGB) intensities are at least 128, otherwise the given pixel is considered to be "dark".

Assume we are given the following table with probabilities:

	Dark	Bright	
Foreground	0.163	0.006	0.169
Background	0.116	0.715	0.831
	0.279	0.721	1

Question: What is the probability of a pixel belongs to the foreground, when it is "dark"?

Let A denote the event that *"the pixel belongs to the foreground"* and let B denote the event that *"dark"*.

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{0.163}{0.279} \approx 0.584 \,.$$

Bayes' rule

Starting with the definition of *conditional probability* $P(B \mid A)$ and multiplying by P(A) we get the **product rule**:

$$P(A \cap B) = P(B \mid A)P(A) \,.$$

Assume that $P(B) \neq 0$. By making use of the *product rule*, we can get

$$P(A \mid B) \triangleq \frac{P(A \cap B)}{P(B)} = \frac{P(B \mid A)P(A)}{P(B)} \,.$$

$P(A \mid B)$ is often called the **posteriori probability**, and $P(B \mid A)$ is called the **likelihood**, and $P(A)$ is called the **prior probability**.

Example: What is the probability that a pixel is "dark", if it belongs to the foreground?
We are given $P(A \mid B) = 0.584$, $P(A) = 0.169$ and $P(B) = 0.279$.

$$P(B \mid A) = \frac{P(A \mid B)P(B)}{P(A)} = \frac{0.584 \cdot 0.279}{0.169} \approx 0.964 \,.$$

Independence

Two events A and B are **independent**, denoted by $A \perp B$, if

$$P(A \mid B) = P(A) \,,$$

equivalently,

$$P(A \cap B) = P(A)P(B) \,.$$

If A and B are **independent**, learning that B happened *does not make A more or less likely to occur.*

Example: Suppose we roll a "fair" die. Let us consider the *events* A denoting *"the die outcome is even"* and B denoting *"the die outcome is either 1 or 2"*.

$P(A) = \frac{1}{2}$ and $P(B) = \frac{1}{3}$. Moreover $A \cap B$ means the event that the outcome is two, that is $P(A \cap B) = \frac{1}{6}$.

$$P(A \cap B) = \frac{1}{6} = \frac{1}{2} \cdot \frac{1}{3} = P(A)P(B) \quad \Rightarrow \quad A \text{ and } B \text{ are } \textit{independent}.$$

Conditional independence

Let A, B and C be *events*. A and B are **conditionally independent** given C, denoted by $A \perp\!\!\!\perp B \mid C$, if

$$P(A \mid C) = P(A \mid B \cap C) \,,$$

equivalently,

$$P(A \cap B \mid C) = P(A \mid C)P(B \mid C) \,.$$

A and B are **conditionally independent** given C means that *once we learned C, learning B gives us no additional information about A.*

Example: The operation of a car's *starter motor* (A) is conditionally independent its *radio* (B) given the *status of the battery* (C).

Summary *

- A **discrete probability space** is a triple $(\Omega, \mathcal{P}(\Omega), P)$, where Ω is countable, $P : \mathcal{P}(\Omega) \to \mathbb{R}$ is σ-additive and for all $A \in \mathcal{P}(\Omega)$ and

$$0 \leqslant P(A) \leqslant 1 \triangleq P(\Omega) \,.$$

- The **conditional probability of A given** B, is the probability that A occurs, given we have observed B.
- If A and B are **independent** ($A \perp B$), learning that B happened does not make A more or less likely to occur.
- A and B are **conditionally independent given** C ($A \perp\!\!\!\perp B \mid C$) means that once we learned C, learning B gives us no additional information about A.

In the **next lecture** we will learn about

- Random variables
- Probability distributions
- Probabilistic Graphical Models

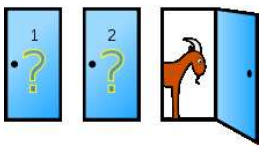
Literature *

1. Marek Capiński and Ekkerhard Kopp. *Measure, Integral and Probability*. Springer, 1998
2. Daphne Koller and Nir Friedman. *Probabilistic Graphical Models: Principles and Techniques*. MIT Press, 2009

A brain teaser *

Suppose you are on a game show and you are given the choice of three doors: Behind one door is a **car**; behind the others, **goats**.

You pick a door, say No. 1, and the *host*, who knows what is behind the doors, opens another door, say No. 3, which has a goat.



Source: wikipedia

He then says to you, "Do you want to pick door No. 2?"

Question: Is it to your advantage to switch your choice?