Computer Vision (ZDO) - Motion analysis

Introduction

Zdeněk Krňoul

Department of Cybernetics
FACULTY OF APPLIED SCIENCES
University of West Bohemia

Content:

- DIFFERENTIAL METHODS
- OPTICAL FLOW
- DETECTION OF FEATURE POINTS
- FREQUENCY APPROACH

Computer vision motion analysis:

- Motion estimation from an image sequence (optical flow);
- Estimation of 3D properties of objects;
- Ego-motion estimation, ie estimation of 3D camera motion to a static scene.
- Necessary step for higher-level processing, which allows you to work with **static and moving position of the observer** and determine **motion parameters**, relative **object** in the image,

Background subtraction methods

- Methods are based on a static background and moving objects in the foreground.
- A moving object has a brightness (or color distribution) different from the background at t; this principle can be summarized in the following formula:

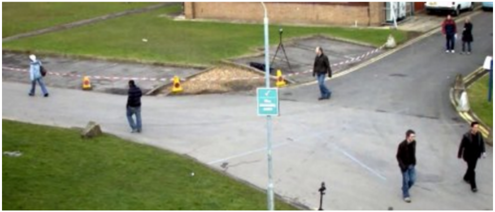
$$F_t(s) = \begin{cases} 1 & \text{for } d(I_{s,t},\ B_s) > \tau, \\ 0 & \text{otherwise.} \end{cases} \qquad (1)$$

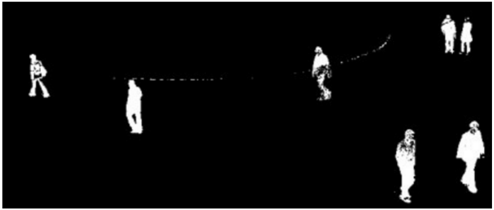
  Where $F_t(s)$ is the foreground at time t at the pixel position s;
- $d(I_{s,t}, B_s)$ indicates the distance between the current image I at time t at pixel position s and background B at pixel position s, τ is the threshold value;
- Background subtraction methods differ in the background model B and distance metric d.

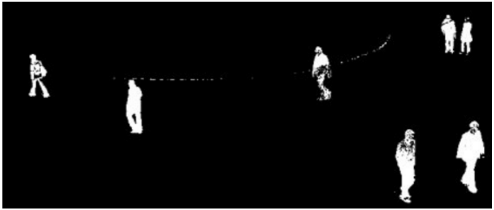
Differential Image

We obtain a binary image d as:

$$d(i,j) = \begin{cases} 0 & \text{for } |f(i,j,t) - f(i,j,t+dt)| < e \\ 1 & \text{otherwise} \end{cases} \quad (2)$$







What causes a value of 1 in the differential image:

- $f(i,j)$ was a background segment at time t and a moving object at time $t+dt$ (or vice versa)
- $f(i,j)$ was a segment of a moving object at t and is an segment of another moving object at $t+dt$
- $f(i,j)$ was at the time t and at time $t+dt$ an element of the same moving object, but in places with different brightness
- Incorrect detected points with a value of 1 will occur due to the presence of noise

Accumulation differential image: we get an intensity image d_{akum}

$$d_{akum}(i,j) = \sum_{t=1}^{T} a_t \cdot |f(i,j,t_0) - f(i,j,t)| \qquad (3)$$

- where $f(i,j,t_0)$ is a reference image
- $f(i,j,t)$ is a sequence of consecutive images
- a_t are weighting coefficients indicating the significance of the individual images of the sequence

note Reference image - an image of the processed scene that contains only stationary objects. If the movement in the scene is continuous, a reference image can be obtained by replacing the areas corresponding to the moving objects with the corresponding areas from other frames.

Adaptive Background Subtraction

- the method solves the problem of determining the reference image (often the background image without moving objects)
- in real conditions, there are related problems with the background as such - e.g. lighting (dimming), a small background change caused by a small movement of the camera (shaking), etc.
- there are several algorithms for adaptive background subtraction

Adaptive background mixture model (y.2001):

- each pixel of the background is a Gaussian mixture model ($K = 3..5$)
- the weights of this mixture model the time with which the given brightness **is in the scene**;
- probable intensities in a given place are background - i.e. they are the longest in the scene and are, therefore, the most stable.

Optical Flow

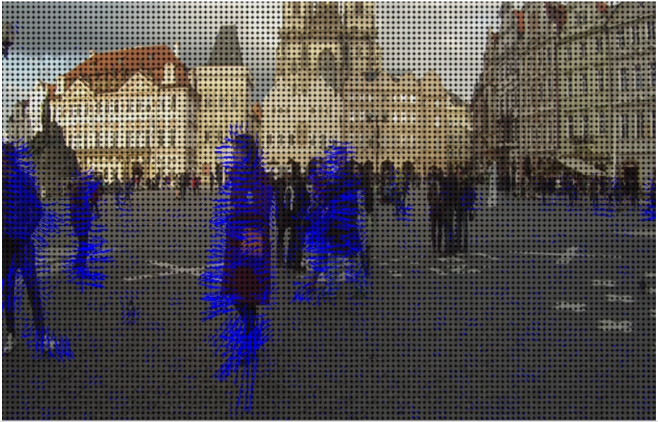
- Analyzes the brightness properties of consecutive images of a given scene over time;
- Sparse optical flow: we analyze the motion of only selected points;
- Dense optical flow: each pixel of the image corresponds to a velocity vector.

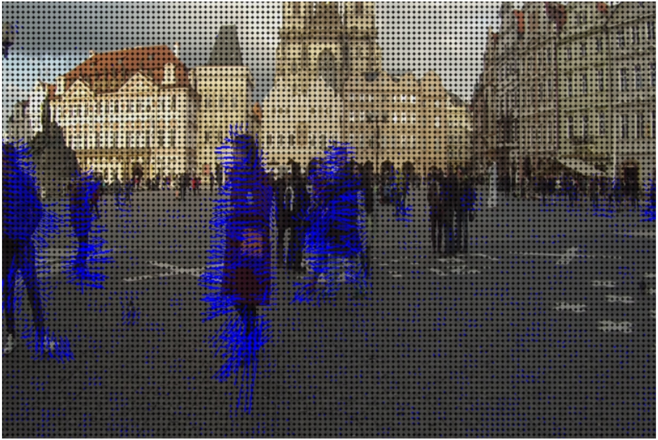
Use of optical flow:

- Object Tracking
- Structure From Motion
- Video Compression
- Video Stabilization

Dense Optical Flow

- Dense optical flow is **array** $2D$ vectors, where each vector shows the displacement of a pixel from a given frame to the next frame;
- is a two-dimensional vector because it determines the direction and magnitude of the velocity at a given position;





Conditions:

1. **the intensity (color) of objects (pixels) does not change in the following image**
2. **adjacent pixels share a similar motion**

Algorithm:

- suppose the intensity function $f(x, y)$ and the pixel intensity $I(x, y, t)$ in a given frame at time t
- the pixel intensity at position (x, y) shifts to the next frame by (dx, dy) for time dt
- suppose the intensity of this pixel is the same, then:

$$I(x, y, t) = I(x + dx, y + dy, t + dt) \tag{4}$$

- then approximation by Taylor series we get the equation of optical flow:

$$f_x u + f_y v + f_t = 0 \tag{5}$$

where

$$\begin{aligned} f_x &= \frac{\partial f}{\partial x} \quad ; \quad f_y = \frac{\partial f}{\partial y} \\ u &= \frac{dx}{dt} \quad ; \quad v = \frac{dy}{dt} \end{aligned} \tag{6}$$

- f_t is the gradient in time and u, v there are unknown values
- The problem is that we have **two unknowns and only one equation**

DEPARTMENT OF CYBERNETICS

- a solutions is the assumption of a common movement of adjacent pixels;
- The Lukas-Kanade method considers a 3x3 block that shares the same motion
- then we have a predetermined system of 9 equations and 2 unknowns;
- the solution is obtained by the least squares algorithm;
- is successful in case of small movement

Modification:

- What about the large movement between adjacent frames?
- The solution is a pyramid representation of these images:
  - gradual resizing images causes large movements to become small movements
  - and small movements are lost
- And we solve the Lukas-Kanade method for each pyramid separately

Sparse Optical Flow

Objective: to solve the problem of **correspondence** of objects at different moments of movement

- The first step is to find the feature points
- a corner detector is often used

Moravec detector: (r.1980)

- is one of the oldest corner detectors

$$g(i,j) = \frac{1}{8} \sum_{k=i-1}^{i+1} \sum_{l=j-1}^{j+1} |f(i,j) - f(k,l)| \qquad (7)$$

- **matching algorithm** searches for correspondence of feature points **in consecutive images**
- the result is a sparse velocity field of these points

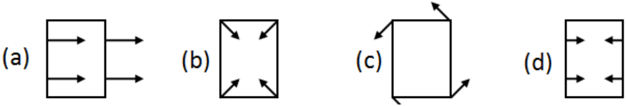
Correspondence:

1. specifying all potential correspondences between pairs of feature points;
2. each pair is evaluated by a certain probability indicating the credibility of their correspondence;
3. probabilities are iteratively refined based on the principle of common motion (across multiple frames);
4. the iteration process ends when there is exactly one corresponding feature point from the next image for each feature point in one image

- we take into account the maximum speed
- consistency of point pairs is also important to find correspondence, i.e. the minimum difference in the speed of movement of these points

Types of movement can be described by a combination of four basic movements:

(a) translational motion in the image plane;
(b) remote translation;
(c) rotation around the view axis;
(d) rotation perpendicular to the view axis.



Phase Correlation

Objective: Estimation of movement between two images

- principle is based on cross-correlation techniques
- we use the frequency spectrum by 2D discrete Fourier transform:
- $G_F = \mathcal{F}\{f(i,j,t)\}$ a $G_B = \mathcal{F}\{f(i,j,t+dt)\}$

$$R = \frac{G_F \circ G_B}{|G_F \circ G_B|} \tag{8}$$

- inverse transformation $r = \mathcal{F}^{-1}\{R\}$

- and the offset is obtained as $(\Delta x \Delta y) = \arg \max\limits_{(x,y)}(R)$
- generally robust to noise, overlaps, etc. (medical, satellite images)
- possible extension by rotation and scale (logarithmic polar coordinates)