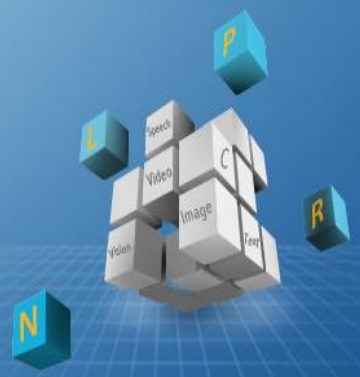
表达与识别

董秋雷
中国科学院自动化研究所
模式识别国家重点实验室



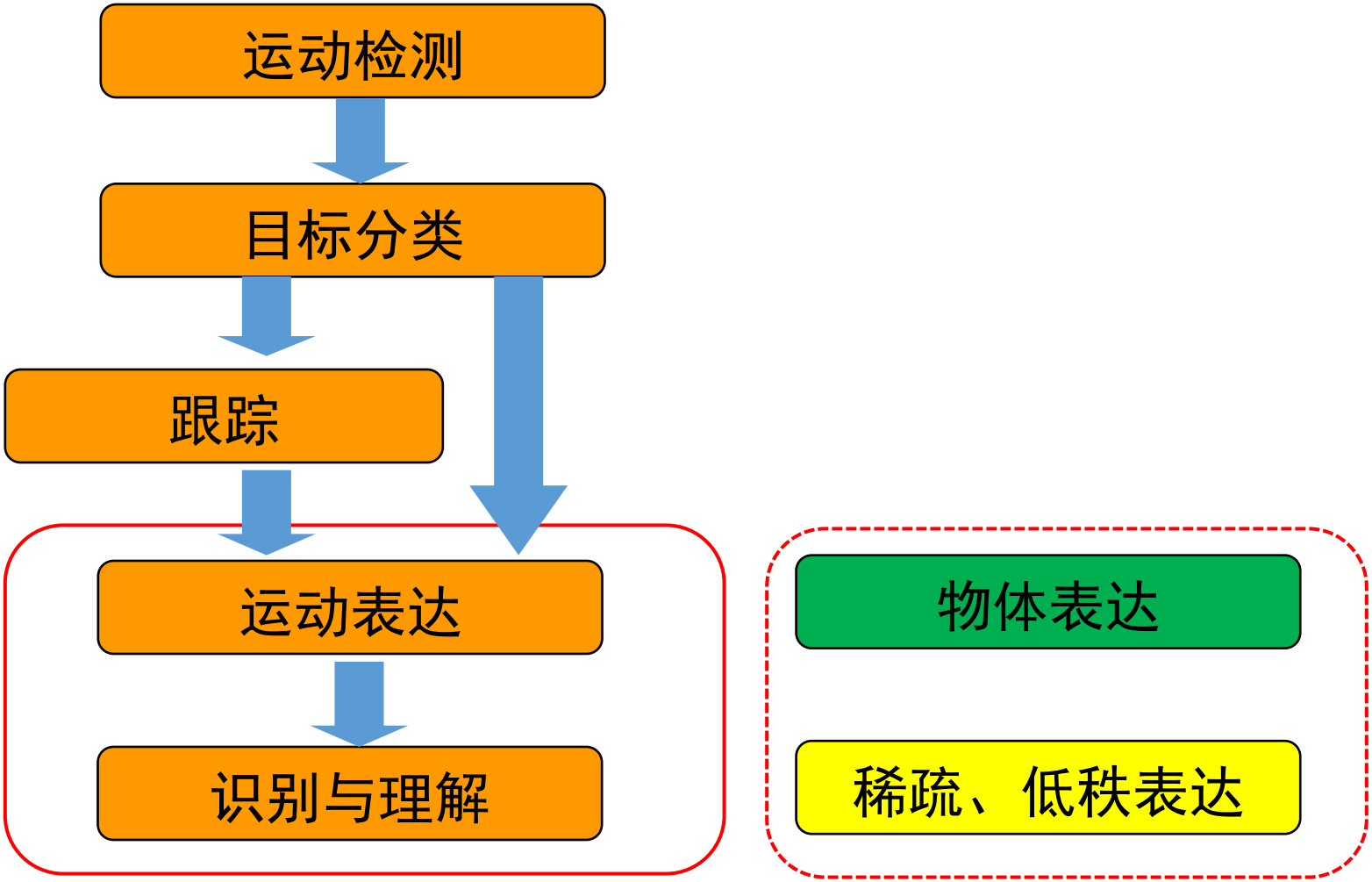
1 背景内容

2 运动表达

3 行为识别

4 小节

运动分析的一般流程



什么是运动表达与行为分析

- 运动表达(Motion representation)？
  - 刻画运动前景的运动模式。
  - 重要性：是运动分析中的中间步骤，是行为理解等高层部分的基础。

- 人的行为分析：
  - 利用计算机视觉技术对视频序列中出现的运动中的人进行检测、跟踪，识别其行为并对其行为进行理解与描述。
  - 应用领域：智能视觉监控、人机交互、增强现实等。

1 背景内容

2 运动表达

3 行为识别

4 小节

① 运动表达

- 运动轨迹
- 时空图表达

② 基于DNN的物体表达

③ 稀疏、低秩表达

- 稀疏表达
- 低秩表达

运动轨迹

- 运动轨迹：通过物体跟踪，可以得到物体特征点的轨迹。
- 能否正确表述物体运动状态的关键：
  - 特征点的选取
  - 轨迹的描述

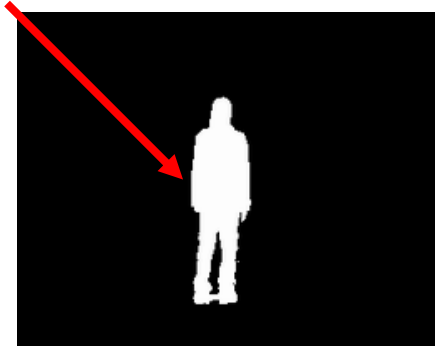
人的运动

车辆的运动

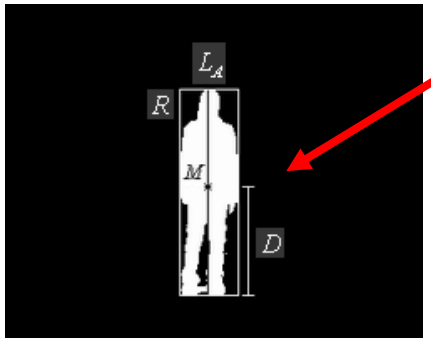
运动轨迹

特征点的选取：

运动前景

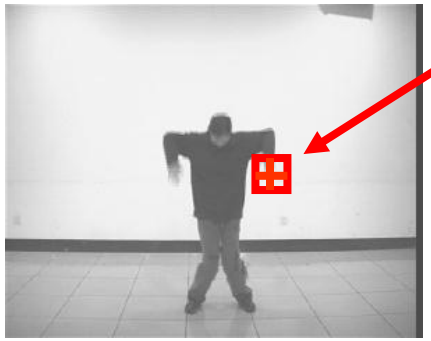


最小包围框



头和脚的特征点

手的质心



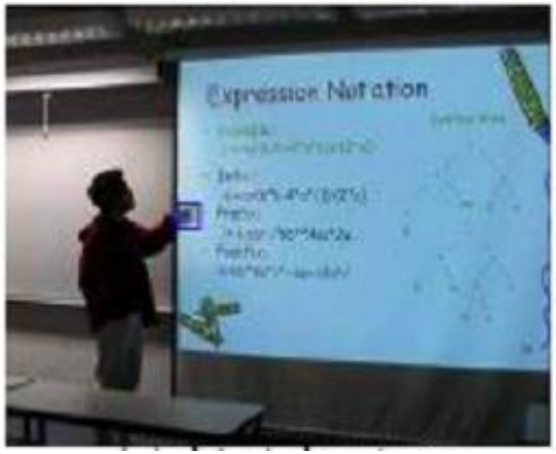
运动轨迹

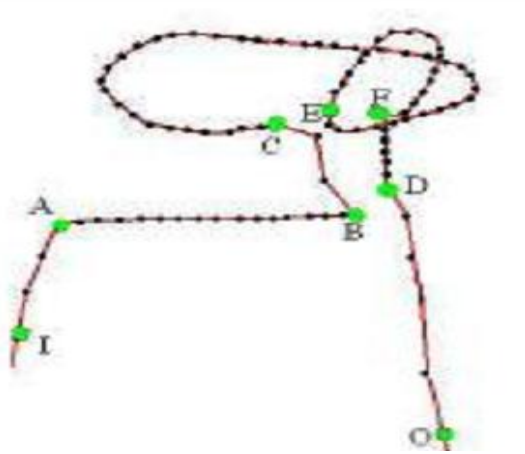
怎样通过特征点集合描述运动轨迹？

1. 直接按照时间顺序连接相邻帧之间的特征点。
2. 将特征点集合拟合成不同的多项式曲线。
3. 其它方法（如主曲线）。

运动轨迹

1. 直接按照时间顺序连接相邻帧之间的特征点。





运动轨迹

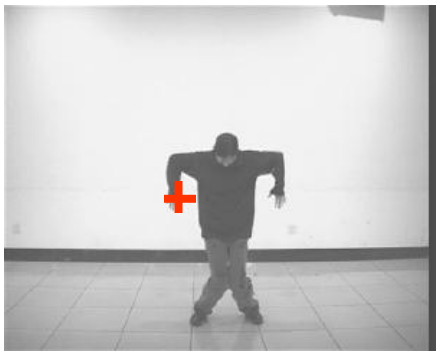
2. 将特征点集合拟合成不同的多项式曲线。如通过最小二乘方法，将一组特征点拟合成一条二次曲线：

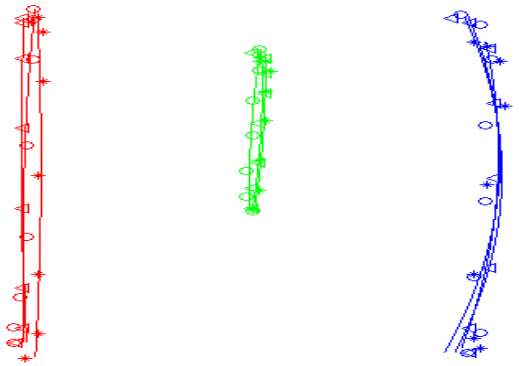
$$ax^2 + 2bxy + cy^2 + 2dx + 2ey + g = 0$$

其中a, b, c, d, e为系数。

运动轨迹



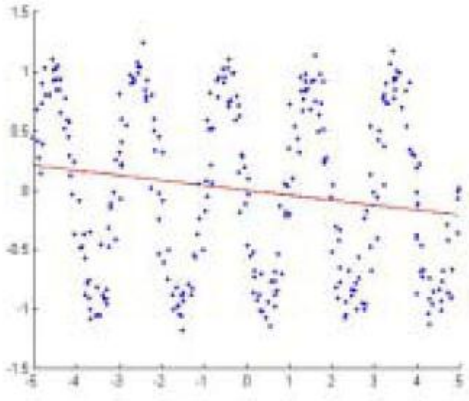


将视频序列中双手和头的质心分别拟合得到的二次曲线

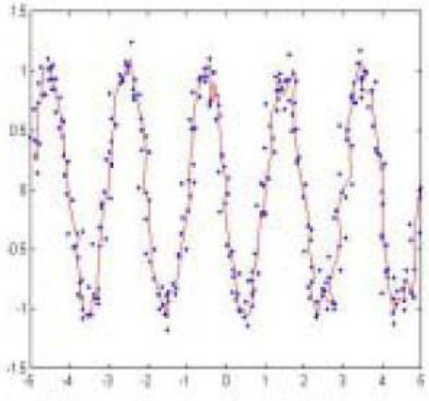
运动轨迹

3. 主曲线[1, 2]：

一条空间曲线，从数据的中部光滑地通过，且不限制于对数据的光滑线性平均，甚至不限制于数据的中部是直线，只使得数据点集合到该曲线的正交距离最小 。



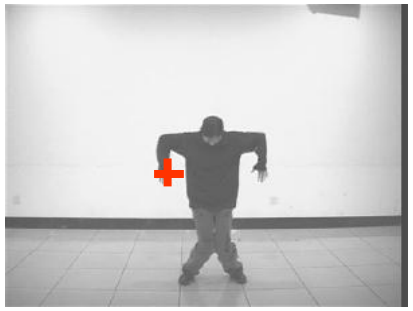
第一主成分线

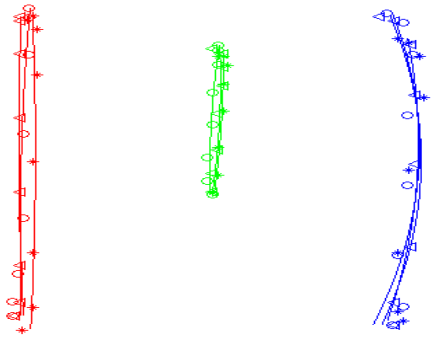


主曲线

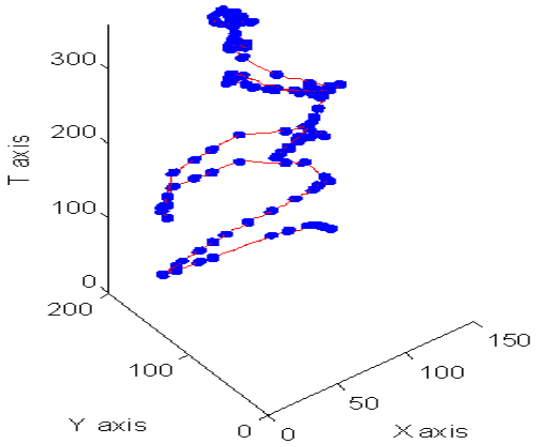
运动轨迹





将视频序列中双手和头的质心分别拟合得到的二次曲线



一段视频序列中特征点集的主曲线

运动轨迹

运动轨迹的应用场合：

- 交通监控，表述车辆、行人行动路线；
- 动作、手势识别，表述运动物体或肢体局部的简单运动；
- 人机接口。

运动轨迹的不足：

- 只能粗略地表述物体全局的运动信息；
- 无法描述运动细节；
- 没有有效地体现时间信息。

时空图表达

- 原理：将图像序列的前景运动信息和时间信息用一张图表述出来。
  - 运动能量图（Motion Energy Image——MEI）
  - 运动历史图（Motion History Image——MHI）
  - 其它“运动图”。

Bobick A., Davis J.: The recognition of human movement using temporal templates. IEEE Trans. PAMI **23**(3), 257–267 (2001)

运动能量图

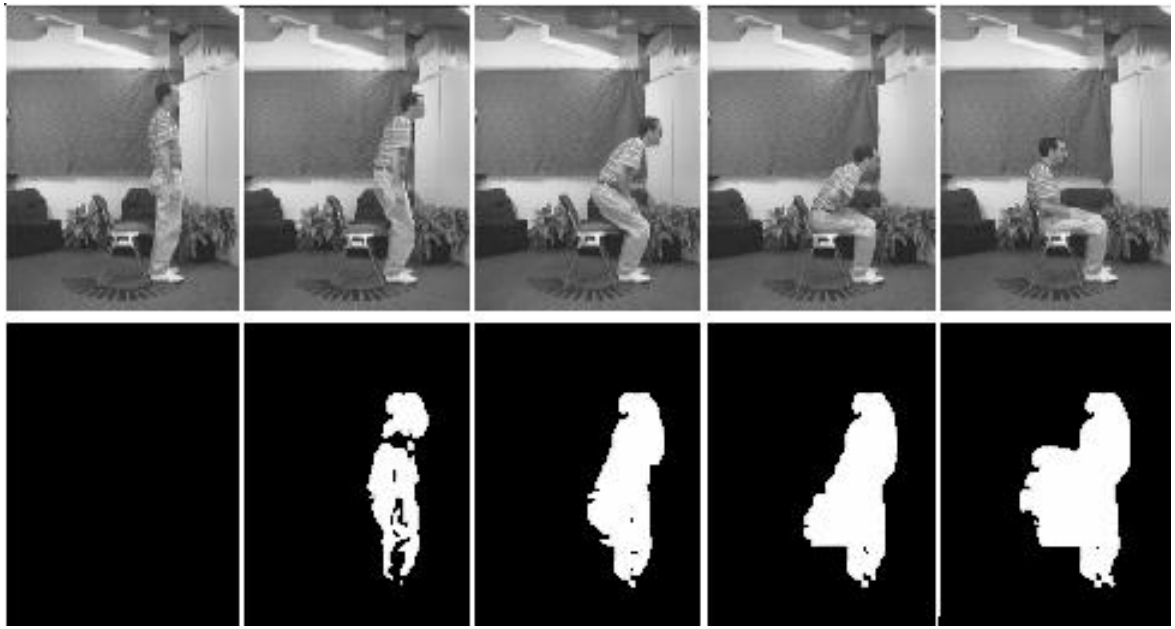
- 前提：通过帧间差分，得到前景的二值图象。
- 运动能量图：将视频序列中所有帧的前景二值化图像求并集。

假设$D(x,y,t)$代表在第t帧与第t－1帧之间差分上的得到的二值化前景，则运动能量图为所有二值化图象的并集：

$$E_\tau(x,y,t)=\bigcup_{i=0}^{\tau-1}D(x,y,t-i)$$

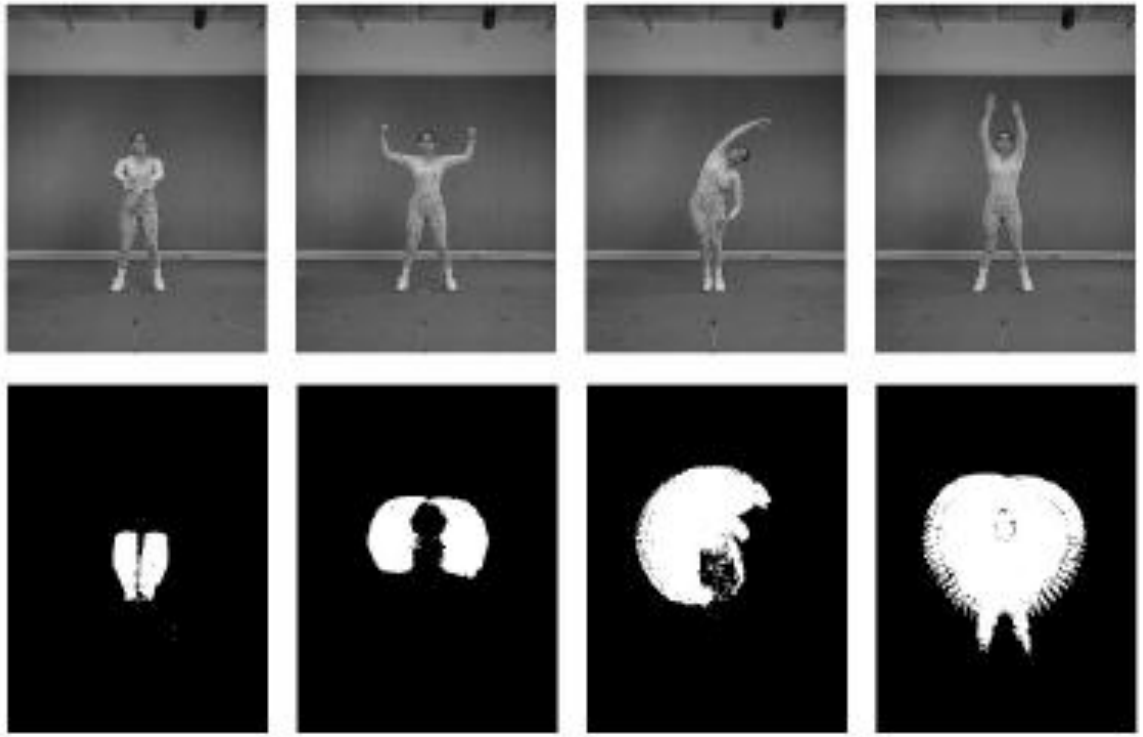
其中参数τ表示一个动作的运动时间。

运动能量图—示例



时间方向

运动能量图—不足



四个对应不同动作的MEI

运动历史图

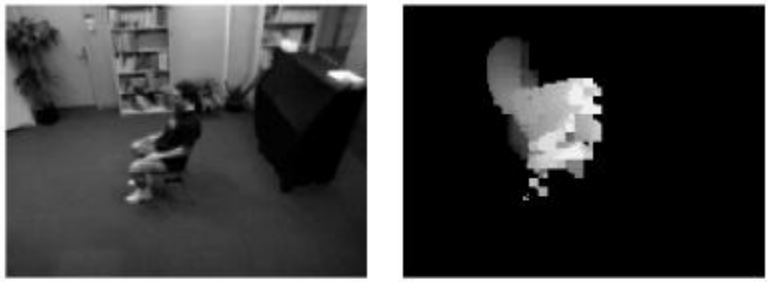
- 前提：通过帧间差分，得到前景的二值图象。
- 运动历史图可以用来表示前景在图像中如何运动的。

假设$D(x,y,t)$代表在第t帧与第t－1帧之间差分上的得到的二值化前景，则运动历史图定义为：

$$H_\tau(x,y,t)=\begin{cases}\tau & D(x,y,t)=1\\ \max(0,H_\tau(x,y,t-1)-1) & otherwise\end{cases}$$

其中参数τ表示一个动作的运动时间。

运动历史图



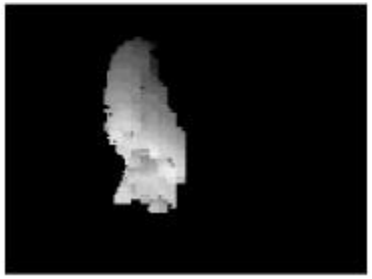
sit-down

sit-down MHI

arms-wave

arms-wave MHI

crouch-down

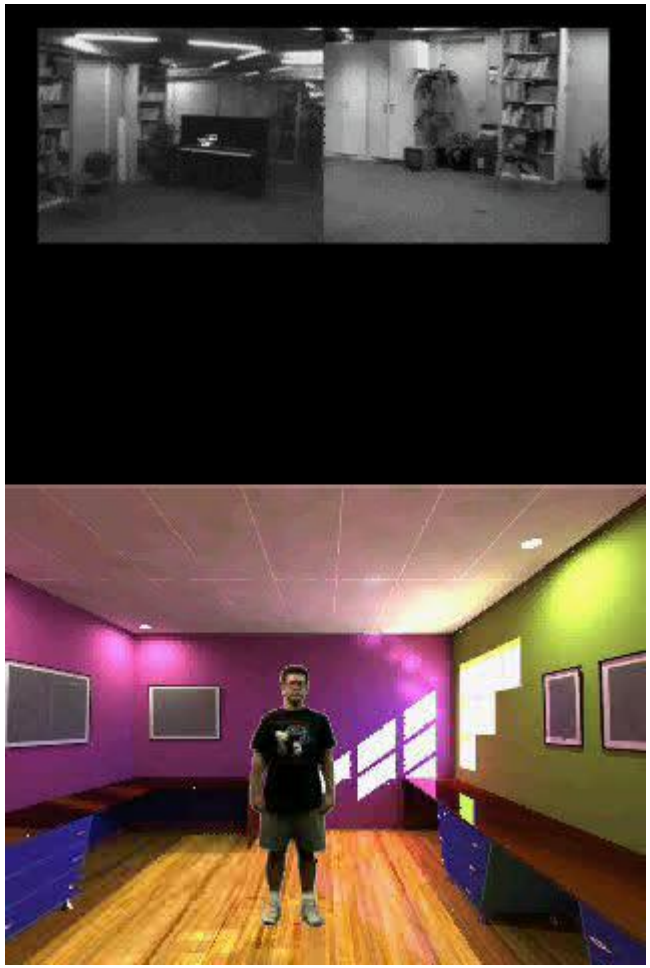


crouch-down MHI

三个对应不同动作的MHI

可见，在运动历史图上，越接近当前帧的运动像素越明亮。

运动历史图—示例



时空图表达

- 主要应用场合：
  - 行为、动作、手势识别；
  - 人机接口。
- 优点：
  - 较好地包含了全局运动、形状、时间信息。
- 不足之处：
  - 缺少局部运动信息，不动有效地区分局部变化的动作；
  - 不动有效地区分速度的变化。

运动表达小节

选择运动表述的原则：具体场景具体分析。

通常情况下，有效的运动表述应具备的特征是：

- 局部运动信息；
- 全局运动信息；
- 时间信息；
- 形状信息等。

① 运动表达

- 运动轨迹
- 时空图表达

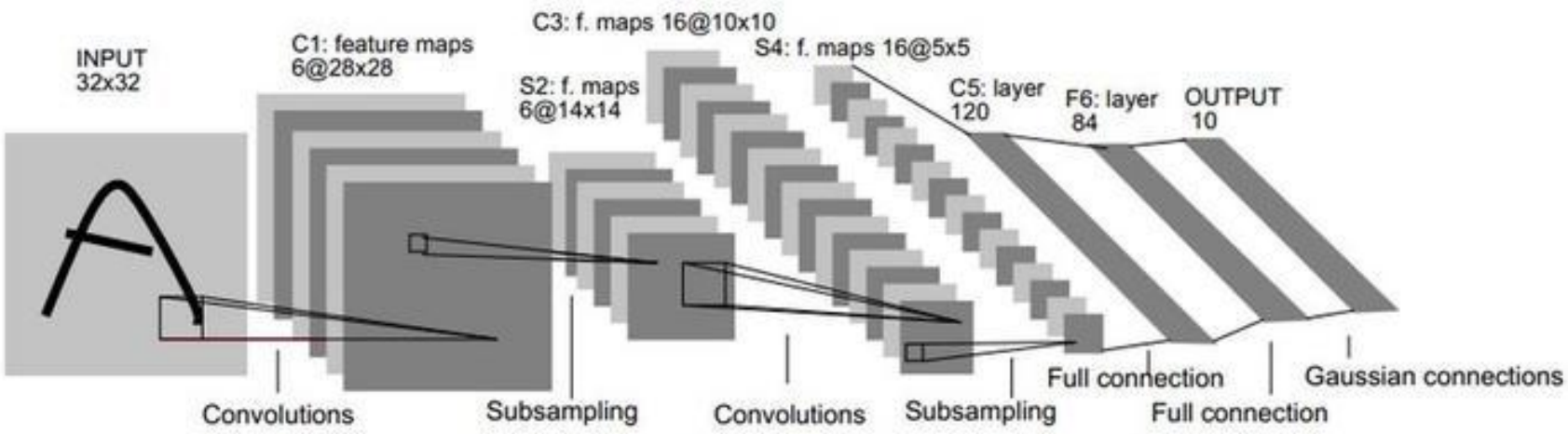
② 基于DNN的物体表达

③ 稀疏、低秩表达

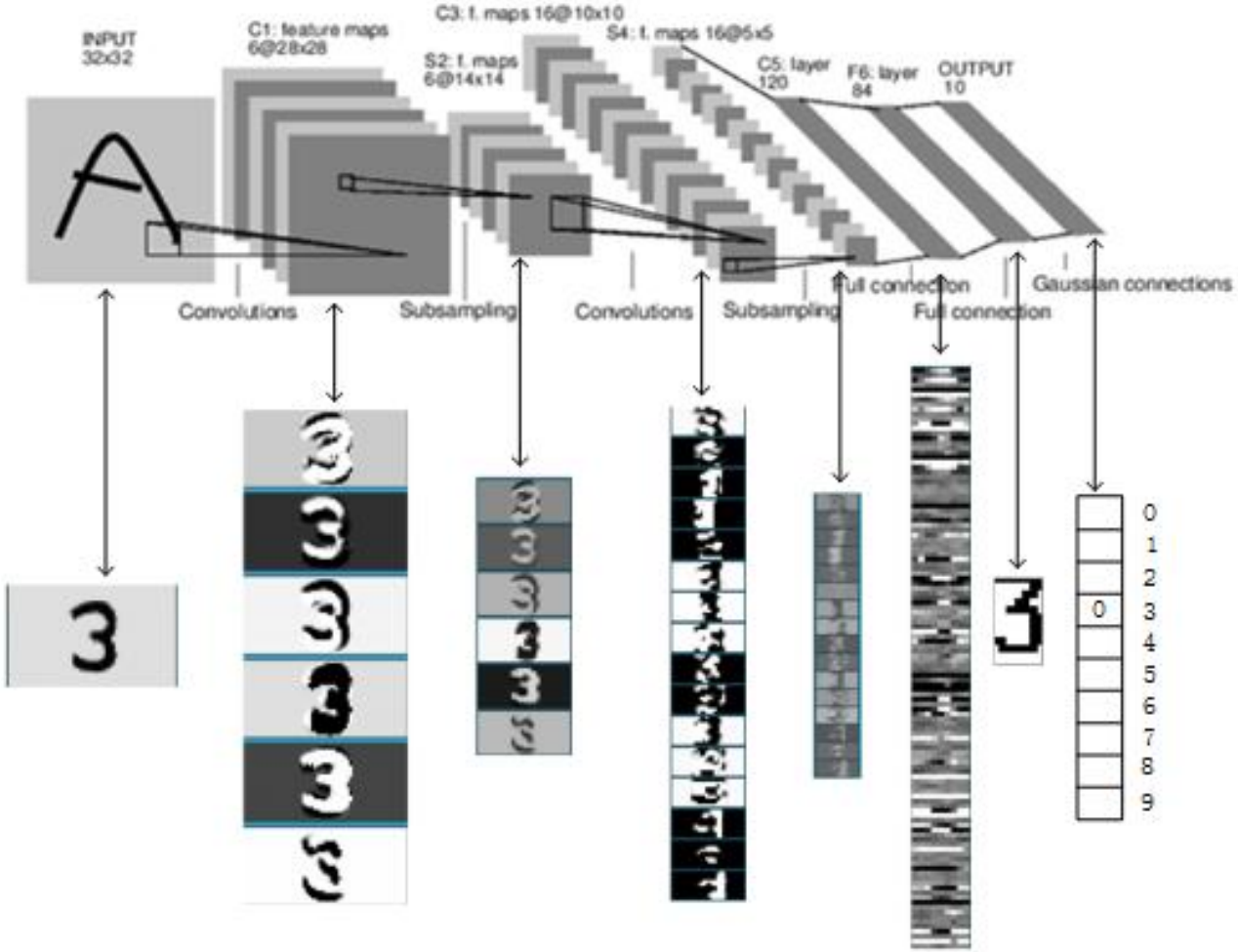
- 稀疏表达
- 低秩表达

基于DNN的物体表达

LeNet：手写数字分类



基于DNN的物体表达


INPUT
32x32
C1: feature maps
6@28x28
S2: f. maps
6@14x14
C3: f. maps 16@10x10
S4: f. maps 16@5x5
C5: layer
120
F6: layer
84
OUTPUT
10
Convolutions
Subsampling
Convolutions
Subsampling
Full connection
Full connection
Gaussian connections
0
1
2
3
4
5
6
7
8
9

基于DNN的物体表达

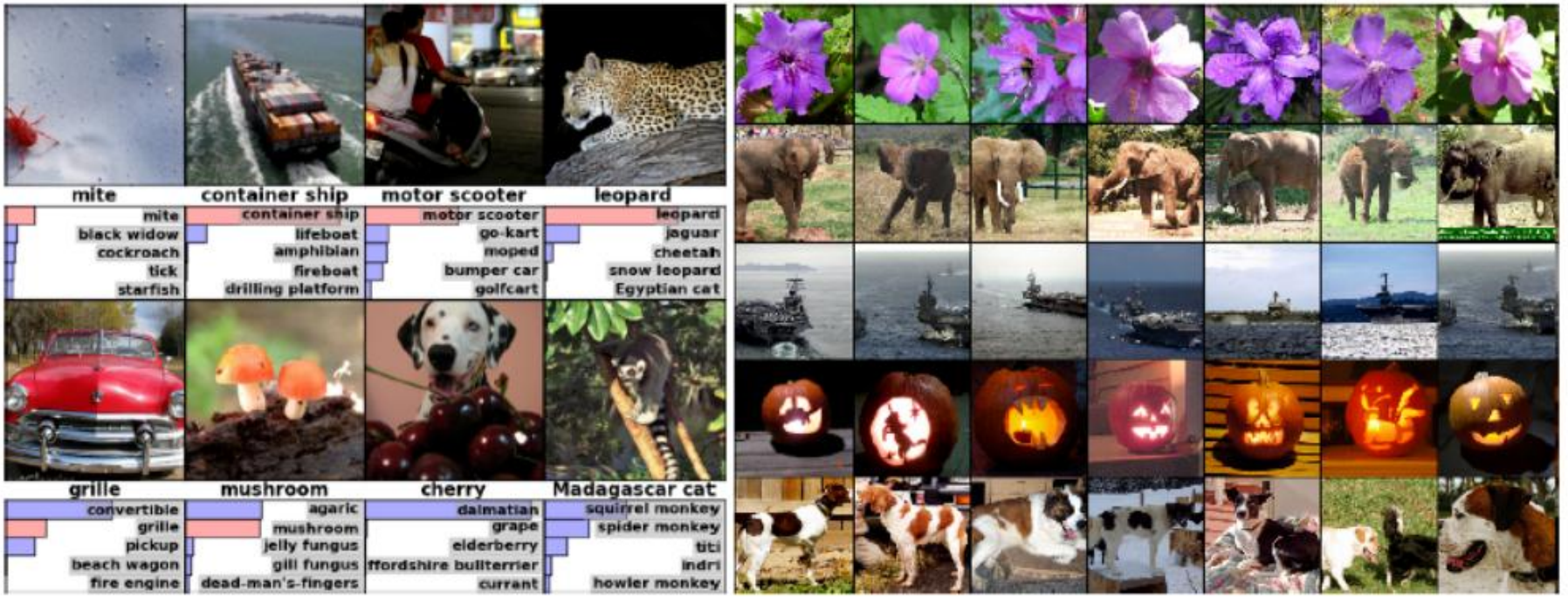
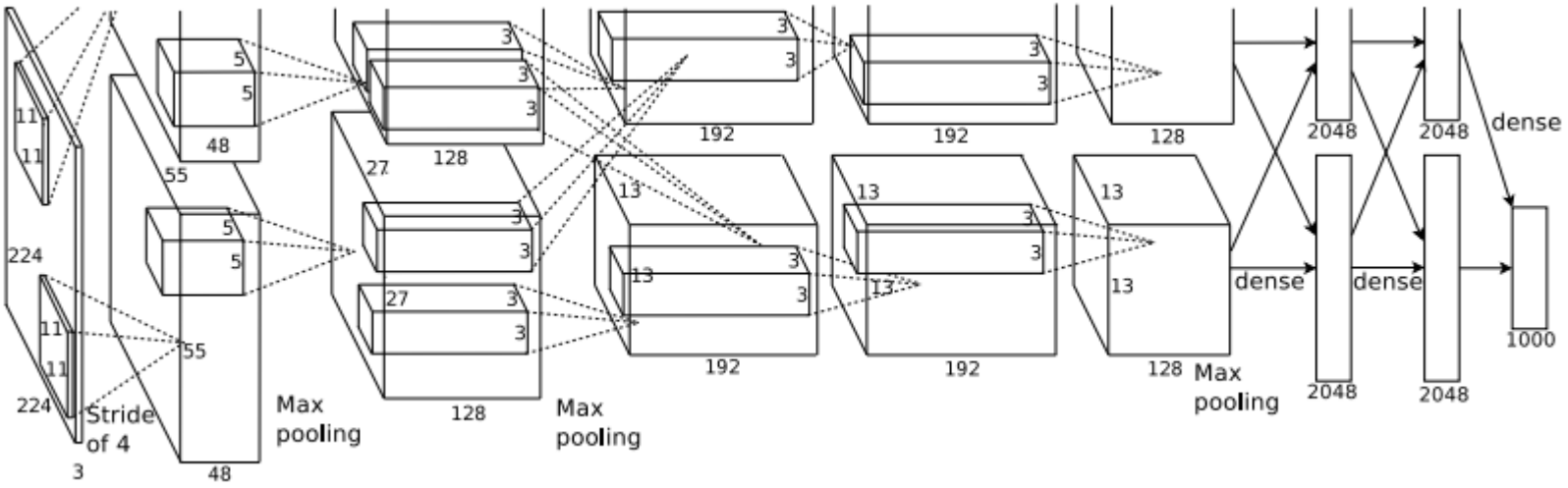


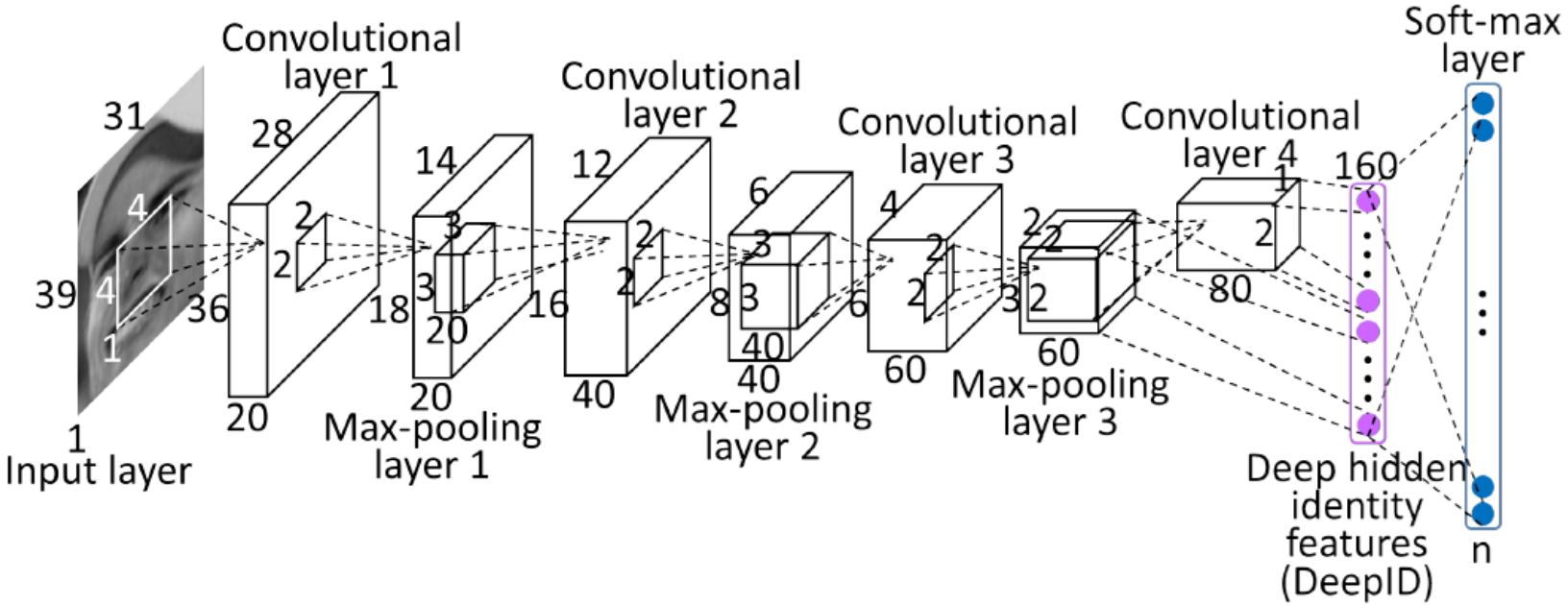
Figure 4: (**Left**) Eight ILSVRC-2010 test images and the five labels considered most probable by our model. The correct label is written under each image, and the probability assigned to the correct label is also shown with a red bar (if it happens to be in the top 5). (**Right**) Five ILSVRC-2010 test images in the first column. The remaining columns show the six training images that produce feature vectors in the last hidden layer with the smallest Euclidean distance from the feature vector for the test image.

基于DNN的物体表达

AlexNet：图像分类

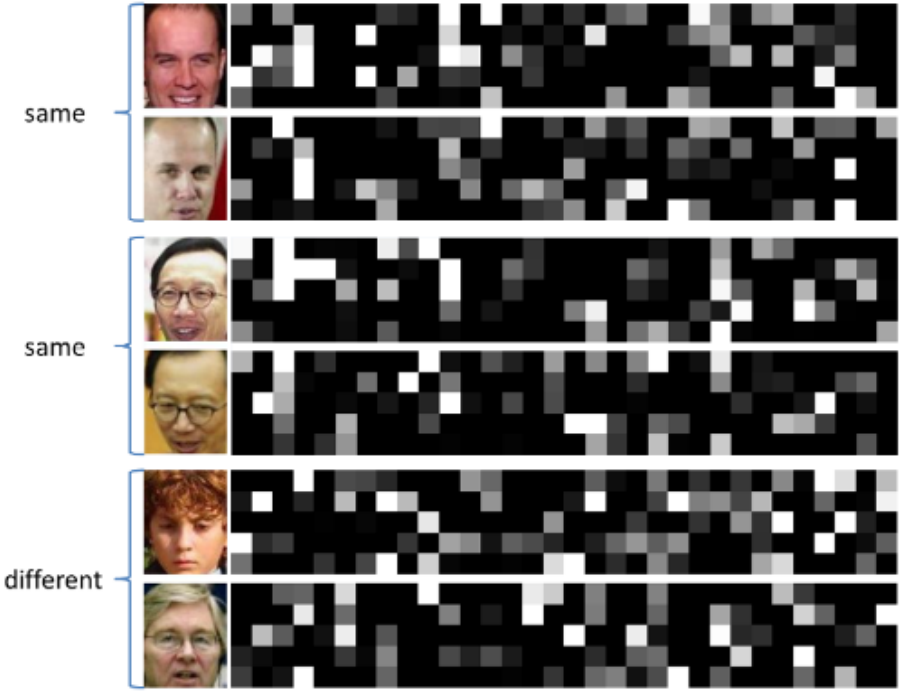


基于DNN的人脸表达



Sun et al. Deep Learning Face Representation from Predicting 10,000 Classes, CVPR 2014

基于DNN的人脸表达



① 运动表达

- 运动轨迹
- 时空图表述

② 基于DNN的物体表达

③ 稀疏、低秩表达

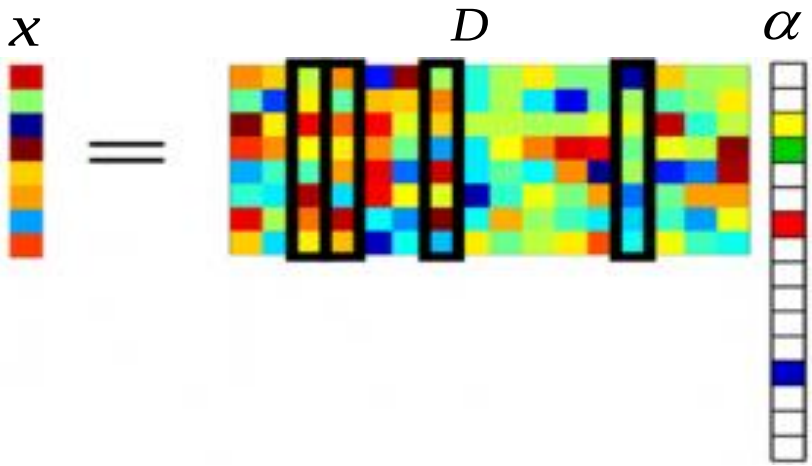
- 稀疏表达（Sparse representation）
- 低秩表达（Low-rank representation）

Sparse representation

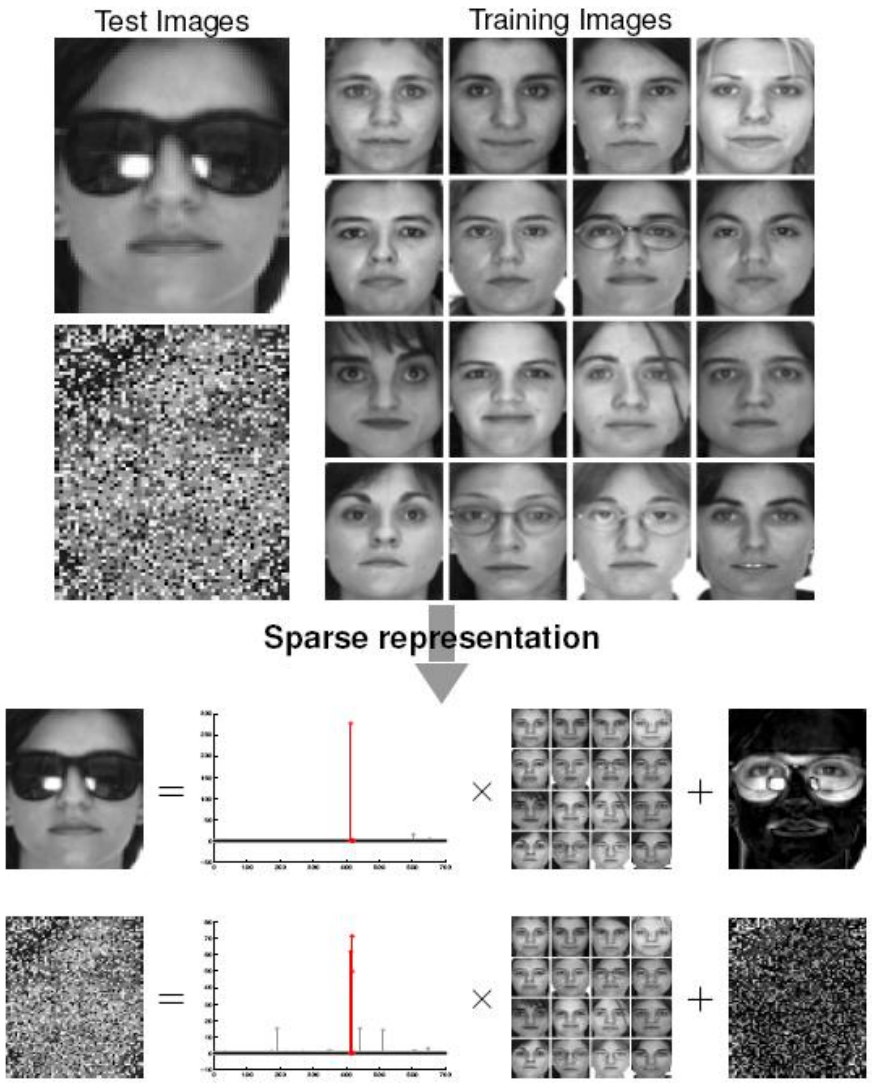
- Original problem: Given $x \in R^m$, $D = [d_1, d_1, ..., d_n] \in R^{m \times n} (m \leq n)$ how to solve

$$x = D\alpha$$

- Sparse representations are the representations that account for most or all information of a signal with a linear combination of a small number of elementary signals.

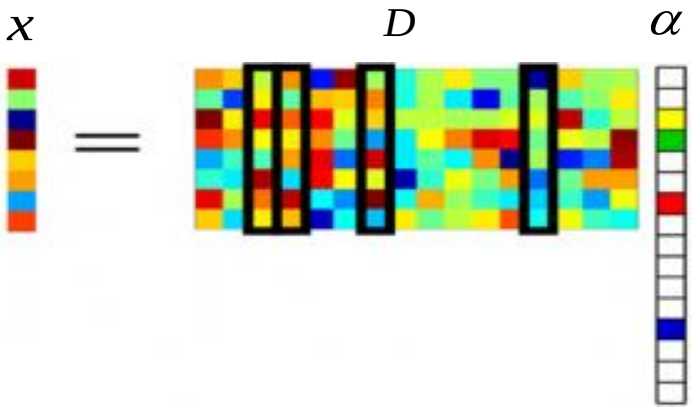


Sparse representation



Sparse representation--Why

- Physiological phenomena: mammalian primary visual cortex (Olshausen and Field,1996)
- Sparsity constraint: The number of the non-zero entries of α are constrained to be as small as possible.



$$\min_{\alpha} \| \alpha \|_0$$

$$s.t. \quad x = D\alpha$$

Olshausen BA, and Field DJ. (1996). "Emergence of Simple-Cell Receptive Field Properties by Learning a Sparse Code for Natural Images." Nature, 381: 607-609

Sparse representation--How

- The constrained problem:

$$\min_{\alpha} \| \alpha \|_0$$

$$s.t. \quad x = D\alpha$$

NP hard

- Donoho(Stanford) and Elad: there exists a unique solution under some condition.

Sparse representation--How

- Candes (Stanford) and Tao(UCLA): Under the RIP(Restricted Isometry Property) condition, the solution to the original L0-norm problem is the same as the one to the corresponding L1-norm problem:

$$\min_{\alpha} \| \alpha \|_0 \quad s.t. \quad x = D\alpha \quad \Rightarrow \quad \min_{\alpha} \| \alpha \|_1 \quad s.t. \quad x = D\alpha$$

Convex!

Sparse representation--How

I. There exists a unique solution to the original L0-norm problem under some condition.

II. Under the RIP(Restricted Isometry Property) condition, the solution to the original L0-norm problem is the same as the one to the corresponding L1-norm problem,

III. The L1-norm problem is a convex problem, which has a unique solution.

Sparse representation

- The L1-norm problem with noise

$$\min_{\alpha} \| \alpha \|_0 \quad s.t. \quad x = D\alpha \qquad \Rightarrow \qquad \min_{\alpha} \| \alpha \|_1 \quad s.t. \| x - D\alpha \|_2^2 \leq \varepsilon$$

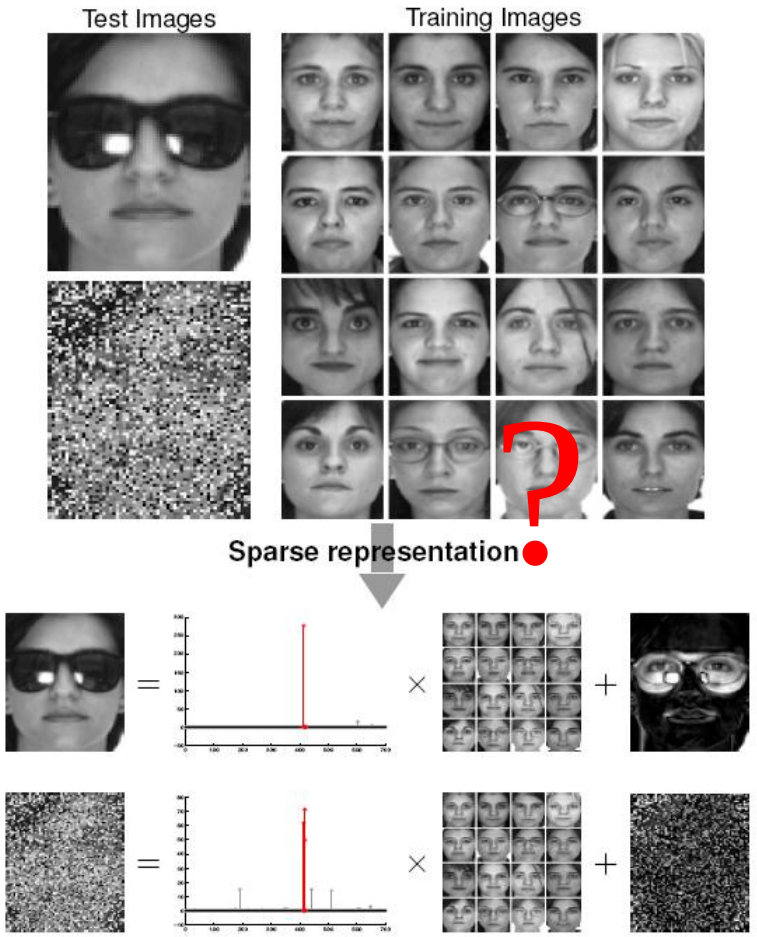
- Algorithms: many

$$\min_{\alpha} \| x - D\alpha \|_2^2 + \lambda \| \alpha \|_1$$

+ *more regularizers*

$$\min_{\alpha} \| \alpha \|_1$$

$$s.t. \| x - D\alpha \|_2^2 \leq \varepsilon$$



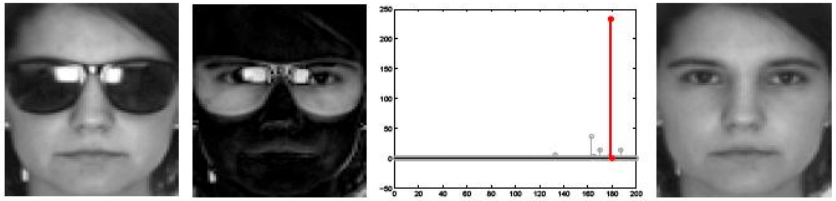
John Wright, Allen Y. Yang, Arvind Ganesh, S. Shankar Sastry, and Yi Ma. Robust Face Recognition via Sparse Representation. PAMI. 31(2) , 210-227, 2009.

Sparse Representation for Recognition

- Assumption: a test sample can be represented by the training samples of the same class.
- Given a set of training samples D(c classes)

$$\hat{\alpha} = \arg\min_{\alpha} \| \alpha \|_1$$

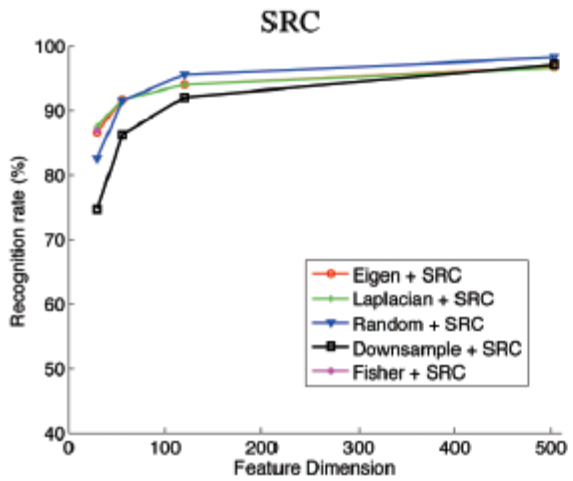
$$s.t. \| x - D\alpha \|_2^2 \leq \varepsilon$$



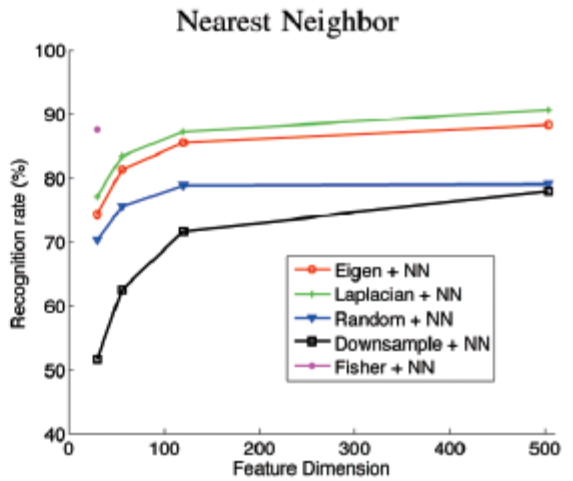
(a) (b) (c) (d)

- The label for x is determined by the minimum reconstruction error:

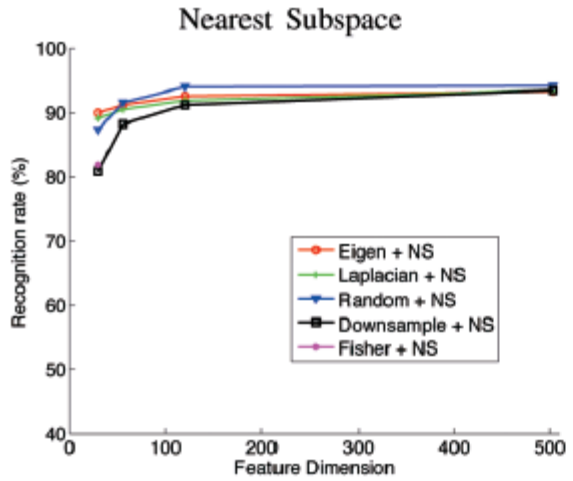
$$\hat{c} = \arg\min_{c} \|\mathbf{D}\delta_c(\hat{\alpha}) - \mathbf{x}\|_2^2 = \arg\min_{c} \|\mathbf{D}_c\hat{\alpha}_c - \mathbf{x}\|_2^2.$$



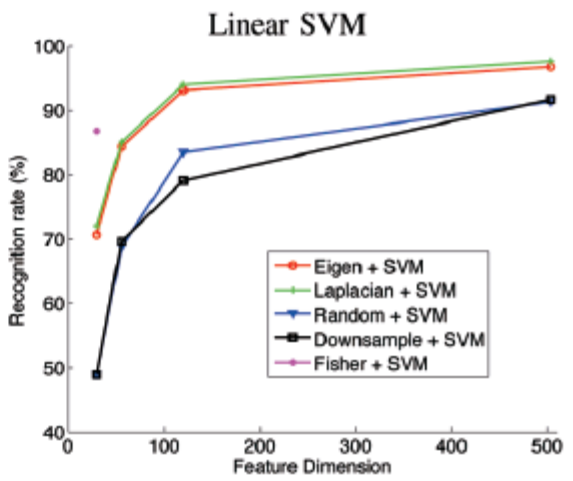
(a)



(b)



(c)



(d)

Extended Yale B database

Comments on sparse representation

- R. Rigamonti, M. Brown and V. Lepetit, Are Sparse Representation Really Relevant for Image Classification? CVPR, 2011.

- Qinfeng Shi, Anders Eriksson, Anton van den Hengel, Chunhua Shen, Is face recognition really a Compressive Sensing problem? CVPR, 2011.

- L. Zhang, M. Yang and X. Feng, Sparse Representation or Collaborative Representation: Which Helps Face Recognition? ICCV, 2011

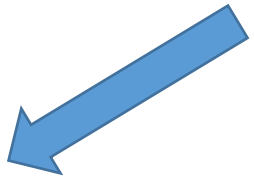
Low-rank representation

Sparse representation

$$\min_{\alpha} \|\alpha\|_0$$
$$s.t. \quad x = D\alpha$$

Low-rank representation

$$\min_{\alpha} \|\alpha\|_{rank}$$
$$s.t. \quad X = D\alpha$$



$$\min_{\alpha} \|\alpha\|_*$$
$$s.t. \quad X = D\alpha$$

$$\min_{\alpha} \|\alpha\|_*$$
$$s.t. \quad X = X\alpha$$

Low-rank representation

$$\min_{\alpha} \|\alpha\|_* \quad s.t. \;\; X = X\alpha$$

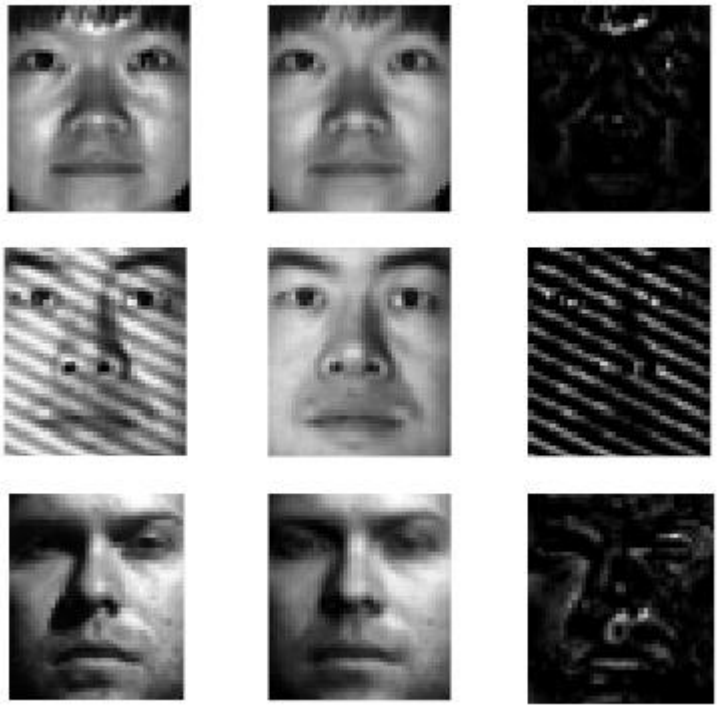
$$\min_{\alpha} \|\alpha\|_* + \|E\| \quad s.t. \;\; X = X\alpha + E$$

$$\min_{\alpha} \|J\|_* + \|E\| \quad s.t. \;\; X = XJ + E \quad \alpha = J$$

Augmented Lagrange Multiplier (ALM)

Low-rank representation

$$X = XZ^* + E^*$$



Liu et al. Robust Subspace Segmentation by Low-Rank Representation, 2010

1 背景内容

2 运动表达

3 行为识别

4 小节

人的行为分析难点

- 人的行为的多样性：
  - 个体行为
  - 人与人之间的交互行为
  - 人和物体之间的交互行为
- 遮挡（人-人；人-物体）情况复杂；
- 人由于穿着的宽松衣物，阴影以及光照变化等因素的影响

人运动的特殊性

- 运动的类型： 刚体运动 vs 非刚体运动
- 人的运动属于非刚体运动中的一个子类：
  - Articulated motion:人体各个部位的运动是刚体运动；而人整体的运动是非刚体运动；

行为识别 — 匹配时变数据

- 行为识别可以看作是时变特征数据的分类问题，即将待识别的行为序列（测试序列）与预先标记好的代表典型行为的参考序列进行匹配。
- 由于人动作执行的差异，匹配行为序列时必须能够处理相似运动模式在空间和时间尺度上轻微的特征变化。

行为识别的两大类方法

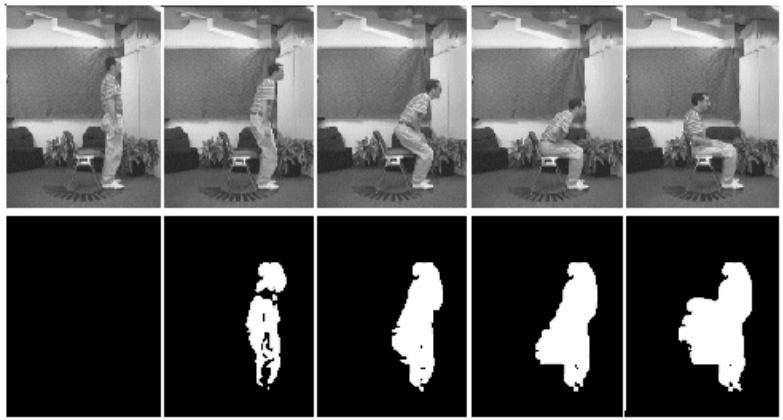
- 基于模板匹配的方法
- 基于状态转移图模型的方法

基于模板匹配的方法

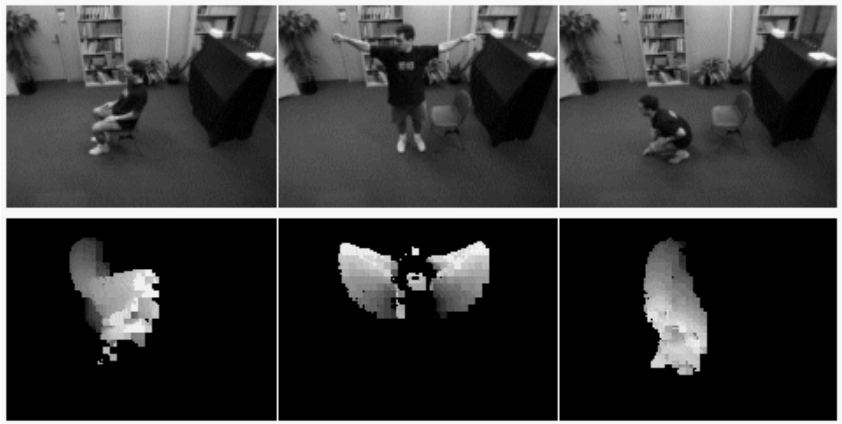
- 基于模板匹配的方法是用输入图像序列提取的特征与在训练阶段预先保存好的模板进行相似性度量，选择与测试序列距离最小的已知模板的所属类别作为被测试序列的识别结果。
  - Temporal Templates (Bobick and Davis PAMI 2001)
  - 动态时间归整 (DTW)

Temporal Templates

- 将图像序列目标的运动信息转化为运动能量图像(MEI)和运动历史图像(MHI)；
- 在图像上提取基于不变矩的运动特征(具有平移、旋转和尺度不变性)，并采用马氏距离度量测试序列和模板之间的相似性。



Motion Energy Image (MEI) : where



Motion History Image (MHI) : how

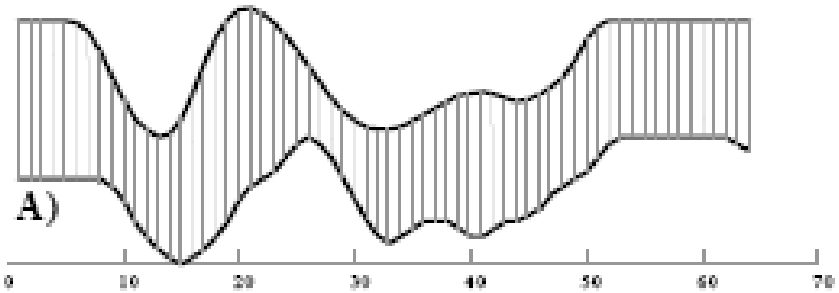
Bobick A and Davis J. **Real-time recognition of activity using temporal templates**.
In: Proc IEEE Workshop on Applications of Computer Vision, Sarasota, Florida, 1996, 39-42.

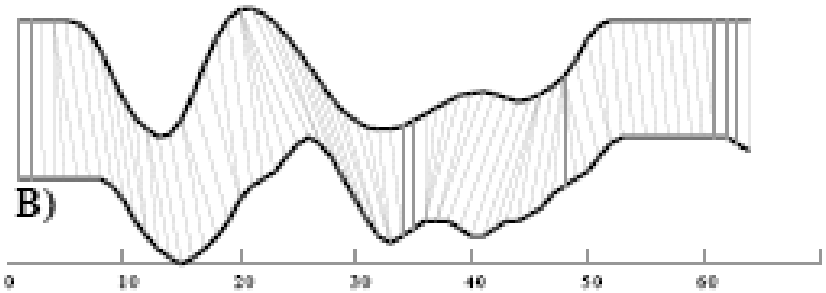
动态时间归整

- 动态时间归整 — Dynamic Time Warping (DTW)
- 是一种时变数据序列匹配方法，常用于微生物学中的DNA匹配、字符串和符号的比较以及语音分析等。

动态时间归整

- 当测试序列模式与参考序列模式的时间尺度不完全一致时：





动态时间归整

- 当测试序列模式与参考序列模式长度不一致时

$$C = \{c_1, c_2, \dots, c_m\} \quad Q = \{q_1, q_2, \dots, q_n\}$$

$$D = \sum_{k=1}^{\min(m,n)} ||c_k - q_k||$$

动态时间归整

假设两个向量C和Q，长度分别为m和n。那么DTW的目标就是要找到一组路径：

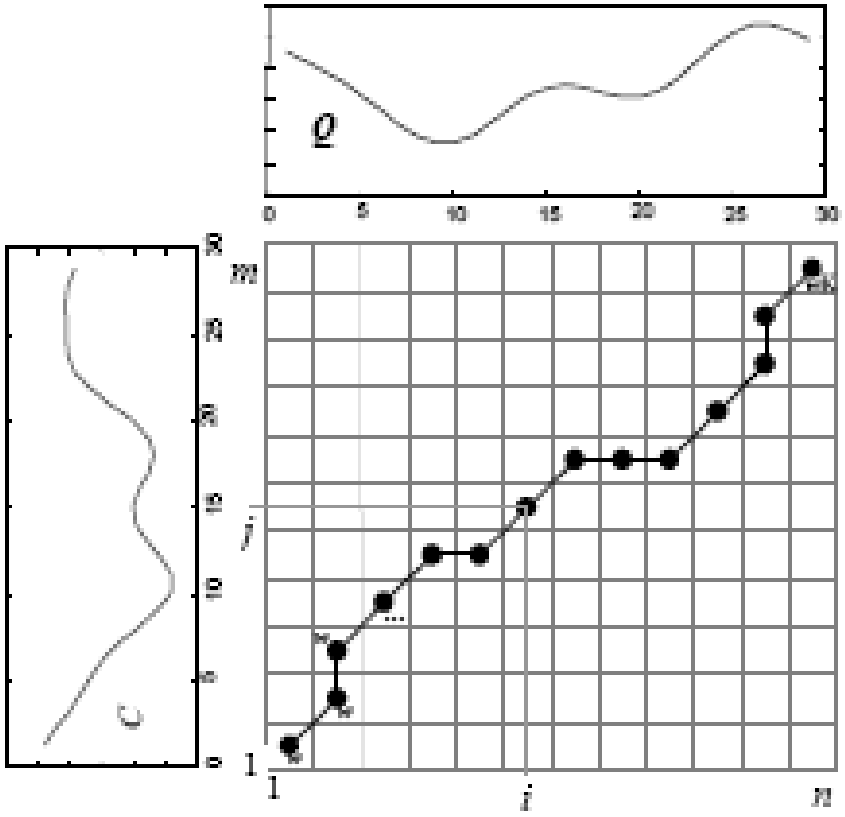
$$W = w_1, w_2, \ldots, w_K \quad \max(m,n) \le K \le m+n-1$$

$$w_k = (c_i, q_j)_k$$

使得经由上述路径的“点对点”对应距离之和为最小：

$$DTW(C,Q) = \min\{\frac{1}{K}\sum_k \|c_i - q_j\|\}$$

动态时间归整(DTW)



* Chu, S., Keogh, E., Hart, D., Pazzani, M. (2002). Iterative Deepening Dynamic Time Warping for Time Series. The Second SIAM International Conference on Data Mining (SDM-02), 2002.

动态时间归整

此路径必须满足以下条件：

1 首尾对齐： $w_1 = (c_1, q_1)$ $w_K = (c_m, q_n)$

2 单调性： $w_k = (a, b), w_{k-1} = (a', b')$

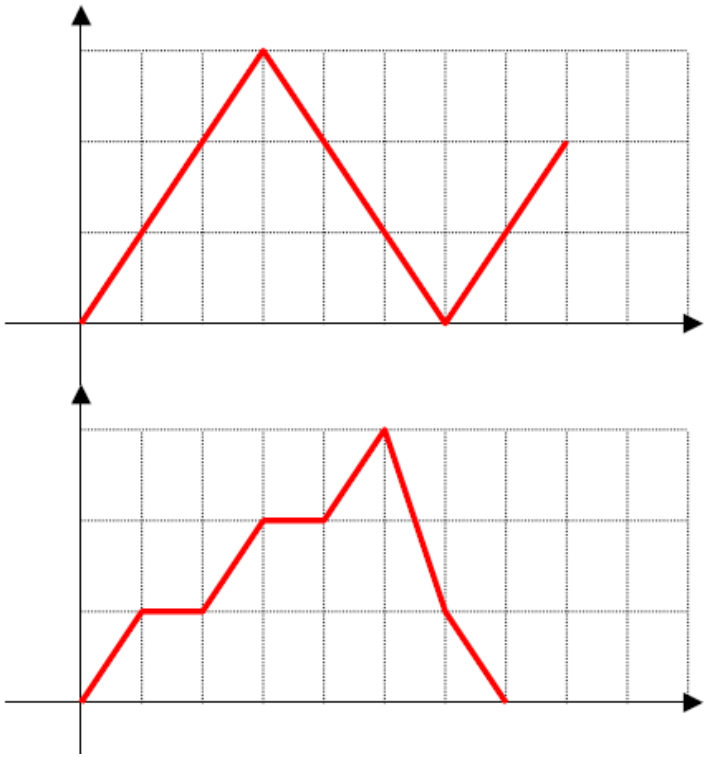
$$0 \le a - a' \le 1, \qquad 0 \le b - b' \le 1$$

示例

Example: Find the alignment path and dissimilarity of the following two sequences, according to the issues explained so far.

$$x[n] = \{0,1,2,3,2,1,0,1,2\}$$

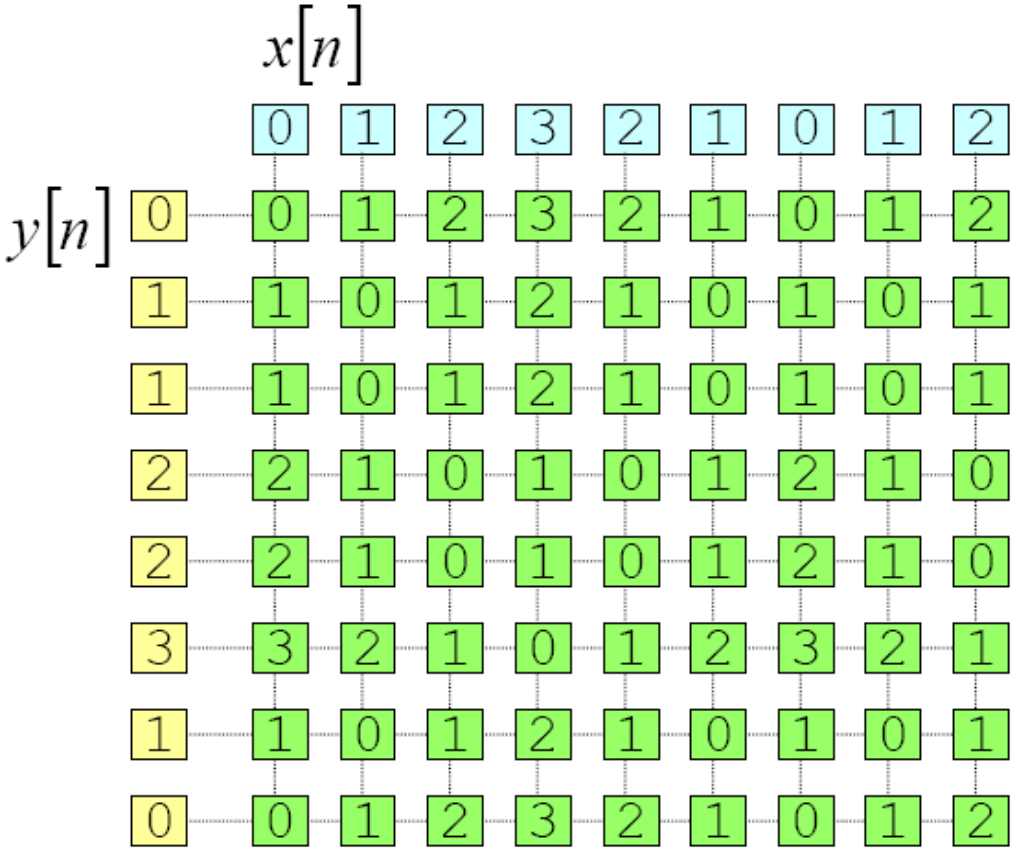
$$y[n] = \{0,1,1,2,2,3,1,0\}$$



示例

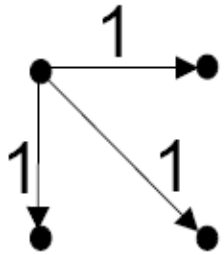
Example.

$$d(i,j) = |x[i] - y[j]|$$



示例

Example.



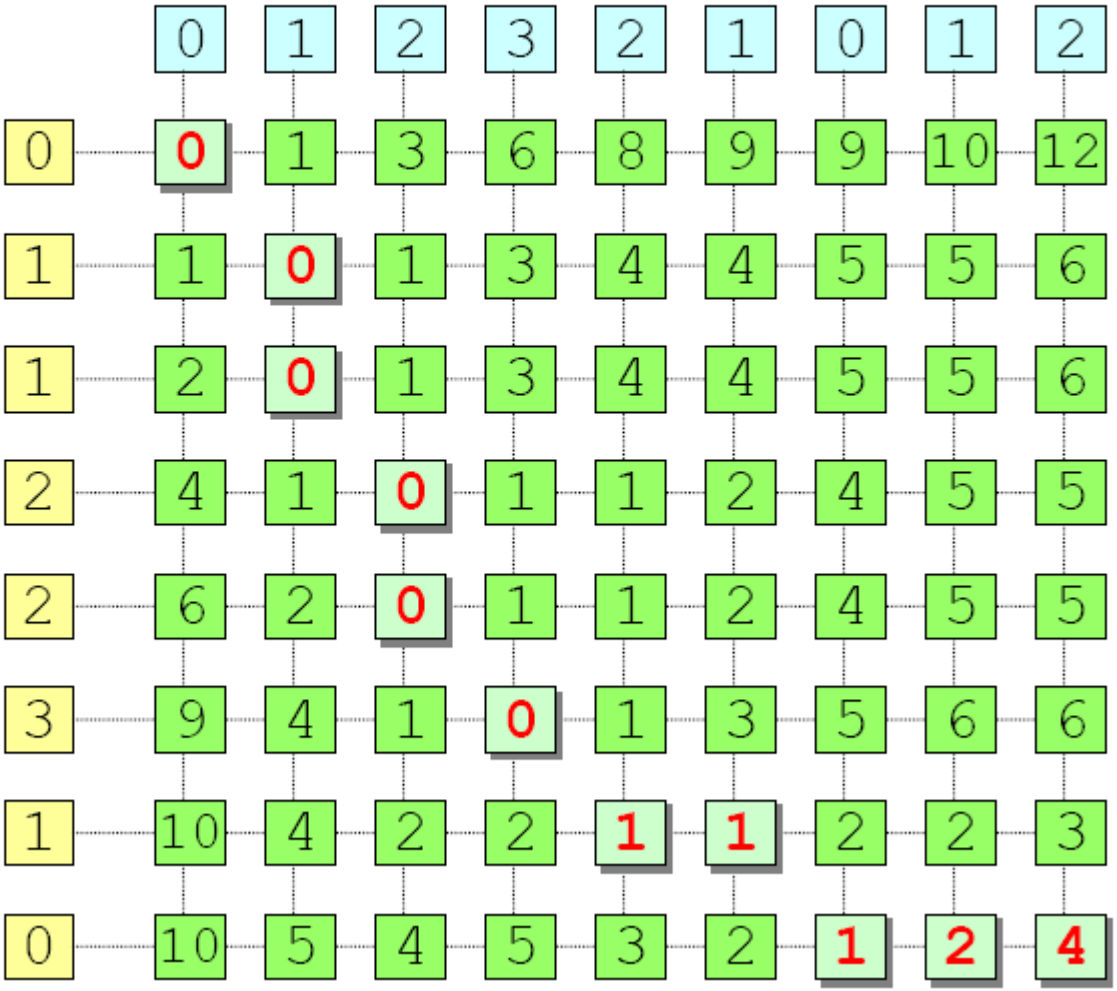
$x[n]$

$y[n]$

$y[n]$	$x[n]$ = 0	1	2	3	2	1	0	1	2
0	0	1	3	6	8	9	9	10	12
1	1	0	1	3	4	4	5	5	6
1	2	0	1	3	4	4	5	5	6
2	4	1	0	1	1	2	4	5	5
2	6	2	0	1	1	2	4	5	5
3	9	4	1	0	1	3	5	6	6
1	10	4	2	2	1	1	2	2	3
0	10	5	4	5	3	2	1	2	4

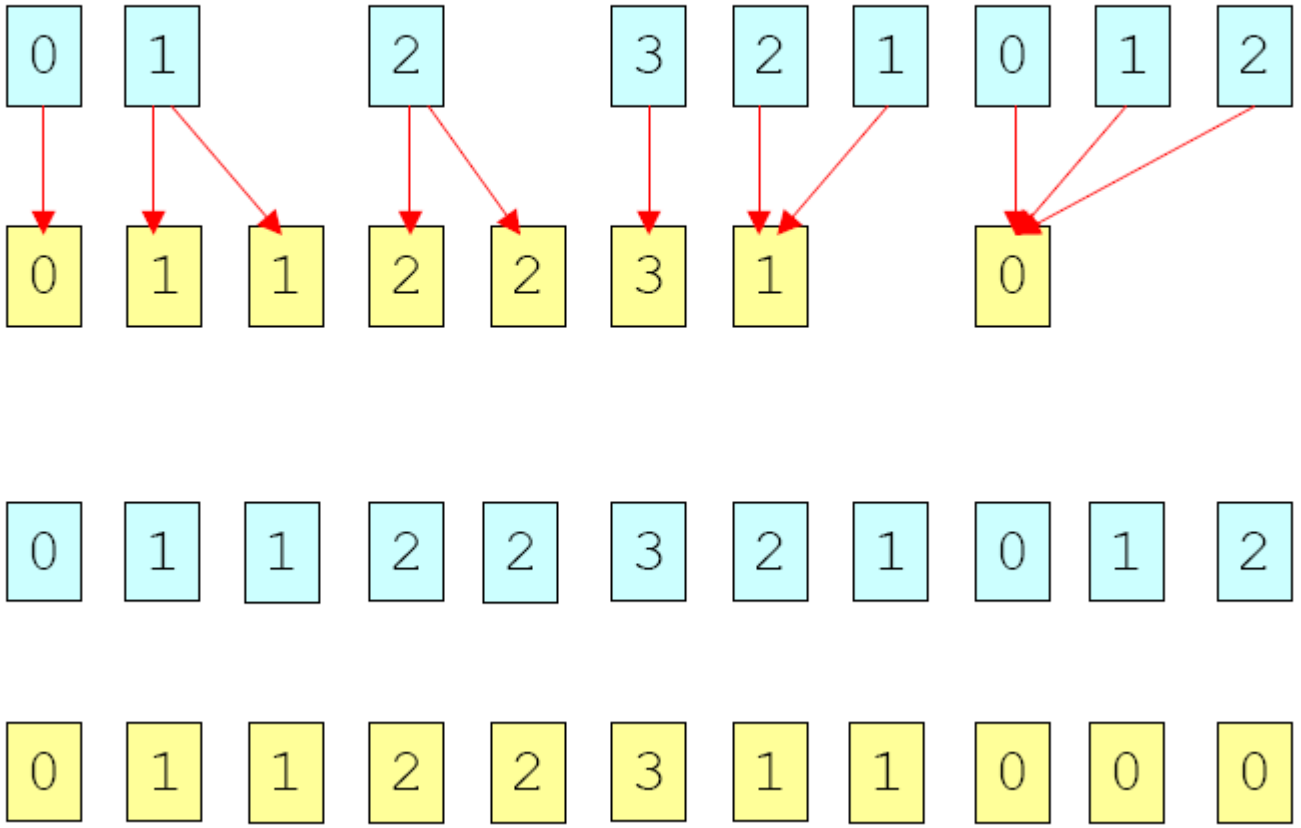
示例

Example. Boundary Constraints.



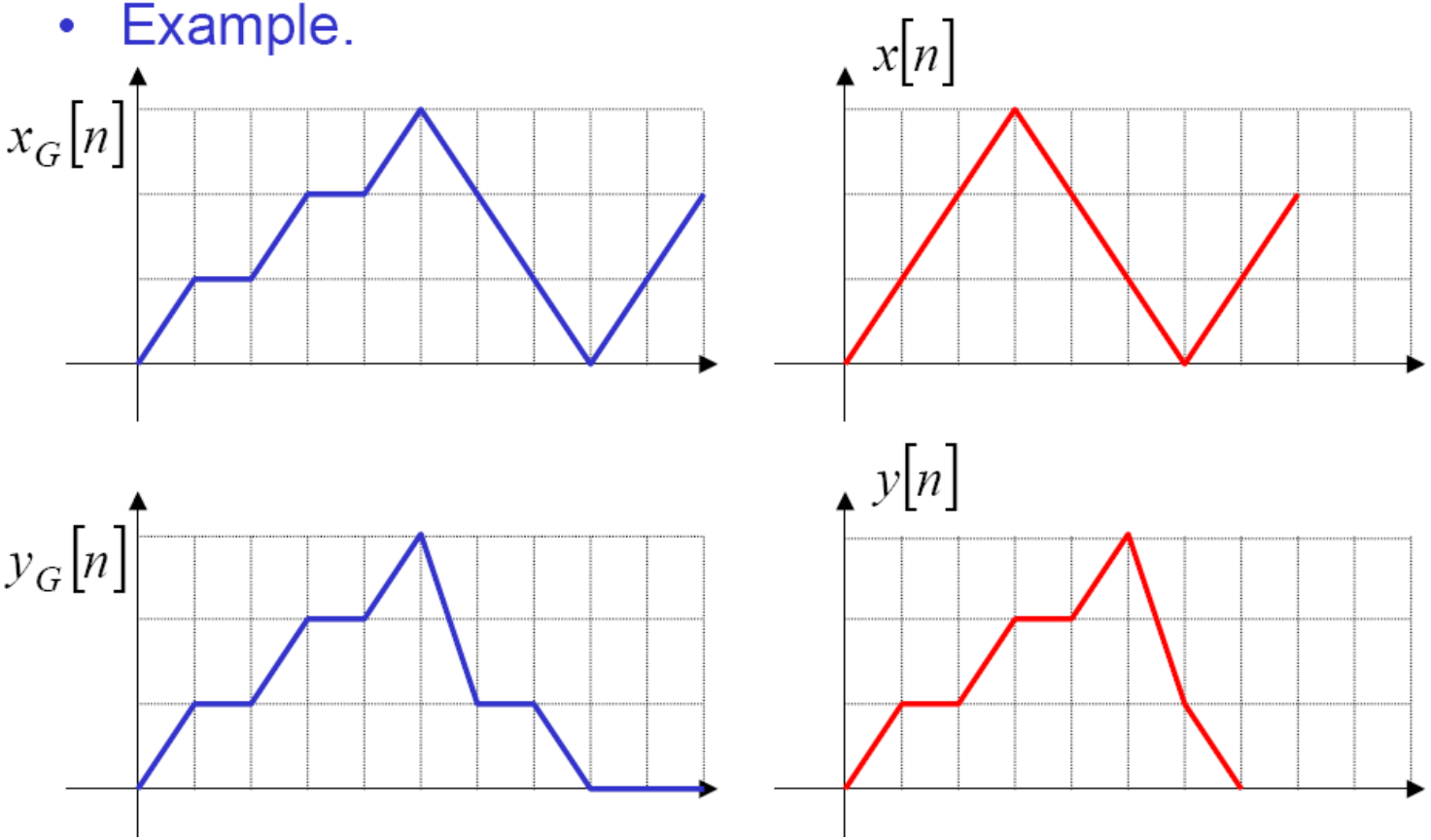
示例

Example.



示例

- Example.



动态时间归整

- 既使测试序列模式与参考序列模式的时间尺度不完全一致，只要时间次序约束存在，DTW就能较好地完成测试序列和参考序列之间的模式匹配。
- 基于动态规划（dynamic programming）思想

基于状态转移图模型的方法

- 基于状态转移图模型的方法定义每个静态姿势作为一个状态，这些状态之间通过某种概率联系起来。任何运动序列可以看作为这些静态姿势的不同状态之间的一次遍历过程，在这些遍历期间计算联合概率，其最大值被选择作为分类行为的标准。
- 常用于行为识别与理解的图模型方法有：
  - 隐马尔可夫及其改进模型
  - 动态贝叶斯网络
  - 人工神经网络
  - 有限状态机
  - 置信网络

Markov, Andrei Andreevich

- 马尔可夫过程：
  - 在已知目前状态（现在）的条件下，它未来的演变（将来）不依赖于它以往的演变（过去）。

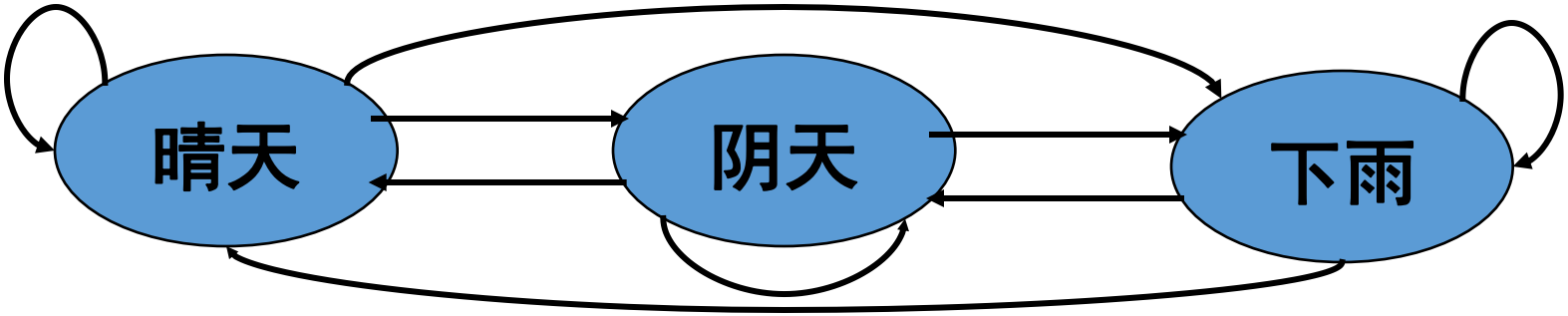
$$X(t+1) = f(X(t))$$

马尔可夫链

- 时间和状态都离散的马尔科夫过程称为马尔科夫链
- 记作{Xn = X(n), n = 0,1,2,…}

  在时间集T1 = {0,1,2,…}上对离散状态的过程相继观察的结果
- X = {q1,...qN}为马尔可夫链的状态空间。
- 条件概率aij =p(Xt+1 = qj |Xt = qi)为马尔可夫链的转移概率

转移概率矩阵



	晴天	阴天	下雨
晴天	0. 50	0. 25	0. 25
阴天	0. 375	0. 25	0. 375
下雨	0. 25	0. 125	0. 625

OMM Examples

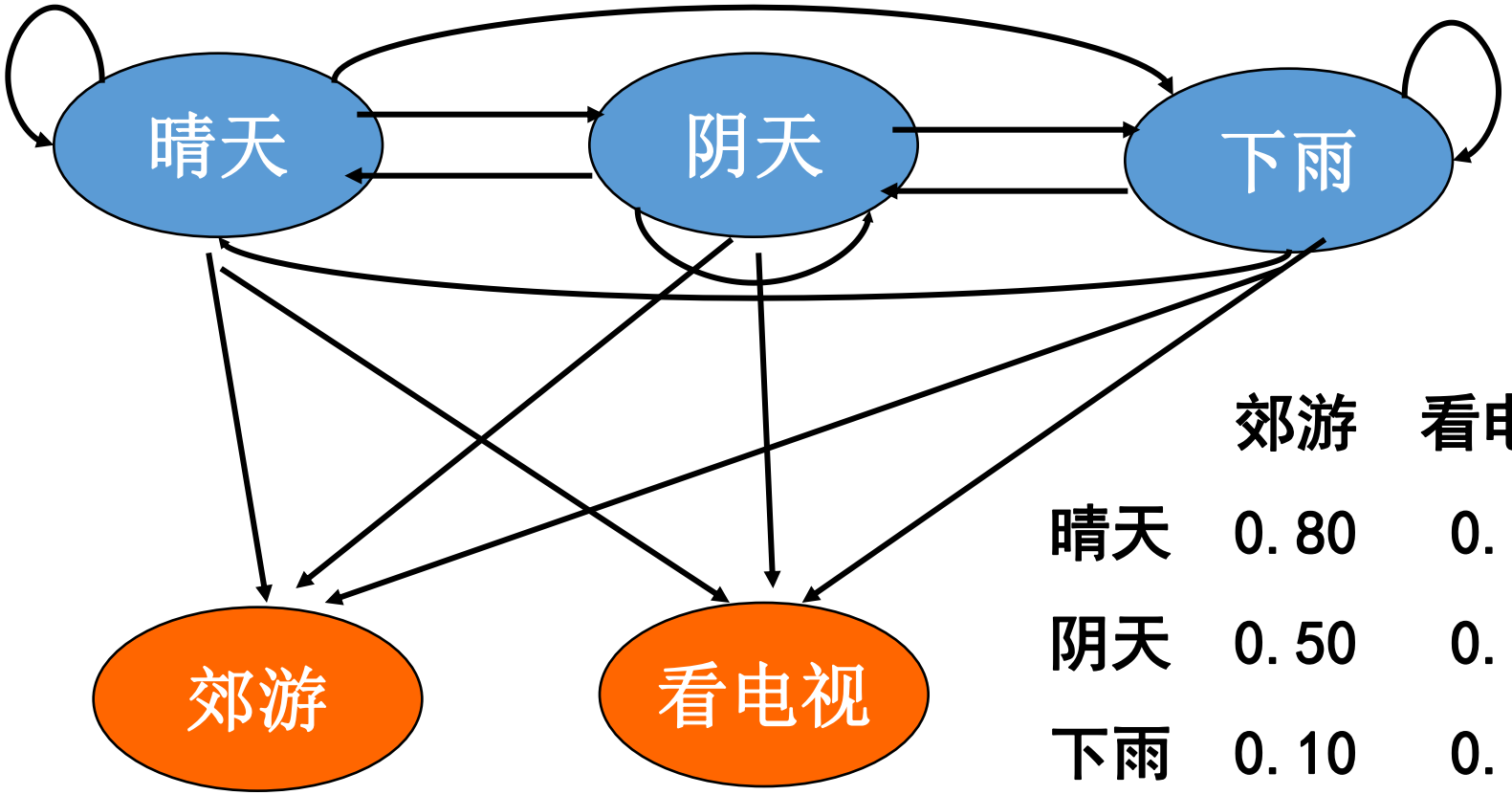
- 第一天天气sunny，接下来7天天气为sun-sun-rain-rain-sun-cloudy-sun...的概率是多少？
- 观察序列O={S3, S3, S3, S1, S1, S3, S2, S3}

	晴天(S3)	阴天(S2)	下雨(S1)
晴天	0.50	0.25	0.25
阴天	0.375	0.25	0.375
下雨	0.25	0.125	0.6

Hidden Markov Model(HMM)

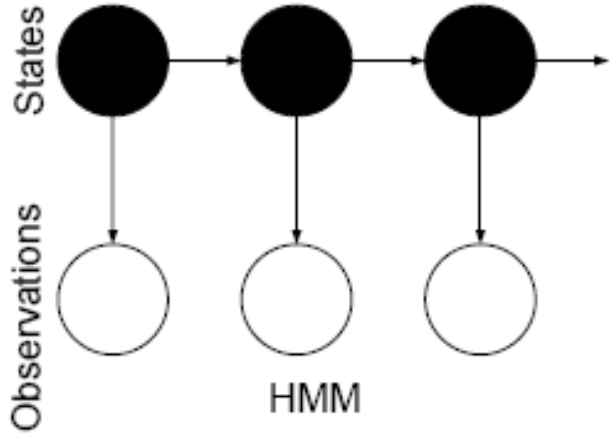
- HMM VS. OMM
  - OMM的每个状态代表一个事件(event),对应一个观察符号(observation)
  - HMM的每个observation对应一个event,每个状态以不同概率发射不同的observation
  - 给定一个观察序列O={$O_1, O_2, ..., O_m$},用OMM能对应还原成对应状态，但用HMM就不行。
  - HMM的状态序列是隐藏于观察序列之下，因此称为hidden

Hidden Markov Model(HMM)



	郊游	看电视
晴天	0.80	0.20
阴天	0.50	0.50
下雨	0.10	0.90

隐马尔可夫模型（Hidden Markov Model）



Markov链
（π, A）

状态序列

$q_1, q_2, \ldots, q_T$

随机过程
（B）

观察值序列

$o_1, o_2, \ldots, o_T$

隐马尔可夫模型

- HMM的状态不可直接观测
- 可观察到的事件与状态并不是一一对应的确定性关系，而是通过一组概率分布相联系
- HMM是一个双重随机过程，两个组成部分：
  - 马尔可夫链：描述状态的转移，用转移概率描述。
  - 一般随机过程：描述状态与观察序列间的关系，用观察概率描述。

隐马尔可夫模型

- 描述：**(X , O, A, B, π)** ← **λ, HMM模型参数**

  $X = \{q1,...q_N\}$： the set of N states;

  $O = \{v1,...,vM\}$： the set of M distinct observations;

  $A = \{a_{ij}\}$， $a_{ij} = p(X_{t+1} = q_j \mid X_t = q_i)$：

  the state transition probability distributions;

  $B = \{b_{ik}\}$， $b_{ik} = p(O_t = v_k \mid X_t = q_i)$：

  the observation probability distributions;

  $\pi = \{\pi_i\}$， $\pi_i = p(X_1 = q_i)$：

  the initial state distribution

隐马尔可夫模型的三个基本问题

- 估值问题
- 解码问题
- 学习问题

隐马尔可夫模型的三个基本问题（1）

- 估值问题：
- 给定训练好了的隐马尔可夫模型参数

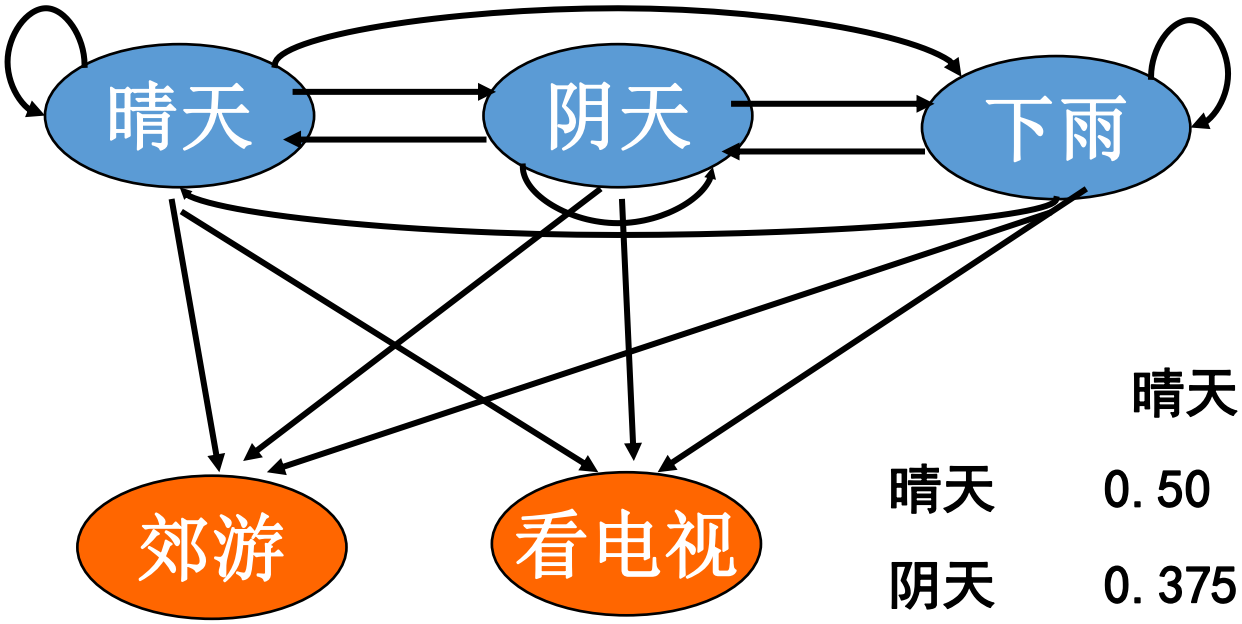
$$\{\lambda_i, i = 1,2,\dots,N\}$$

- 给定待识别的样本序列o

$$\arg\max_i\{P(o|\lambda_i)\}$$

L. R. Rabiner, A tutorial on hidden Markov models and selected applications in speech recognition, Proceedings of the IEEE, vol. 77, no. 2, pp. 257-286,1989.

转移概率矩阵



	郊游	看电视
晴天	0. 80	0. 20
阴天	0. 50	0. 50
下雨	0. 10	0. 90

	晴天	阴天	下雨
晴天	0. 50	0. 25	0. 25
阴天	0. 375	0. 25	0. 375
下雨	0. 25	0. 125	0. 625

清明节： 郊游 看电视 看电视

?

数学表达

给定一个固定的状态序列S=(q1，q2，q3...)

$$P(O|S,\lambda) = \prod_{t=1}^{T} P(O_t|q_t,\lambda) = b_{q_1}(O_1)b_{q_2}(O_2)\cdots b_{q_T}(O_T)$$

$b_{q_t}(O_t)$表示在q_t状态下观测到O_t的概率

$$P(O|\lambda) = \sum_{S} P(O|S,\lambda)\ P(S|\lambda)$$

前向算法（Forward Algorithm）

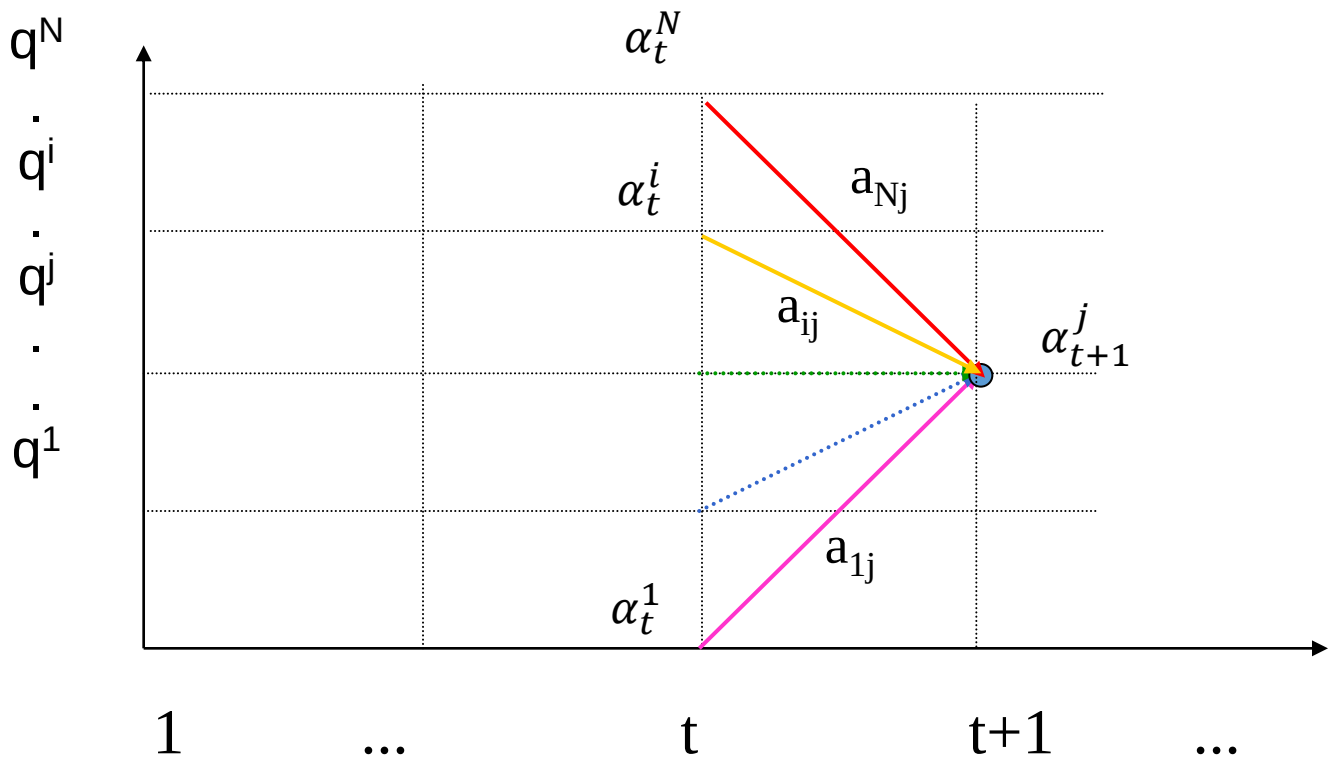
- 前向变量

$$\alpha_t(i) = P(O_1, O_2, \cdots, O_t, q_t = \theta_i | \lambda)$$

- 则有：

$$\alpha_1(i) = \pi_i b_i(O_1)$$

$$\alpha_{t+1}(j) = [\sum_{i=1}^{N} \alpha_t(i) a_{ij}] b_j(O_{t+1})$$


q^N
.
q^i
.
q^j
.
.
q^1
α_t^N
α_t^i
a_{Nj}
a_{ij}
α_{t+1}^j
a_{1j}
α_t^1
1
...
t
t+1
...

前向算法（Forward Algorithm）

算法步骤

I. 初始化：$\alpha_1(i) = \pi_i b_i(o_1)$

II. 迭代：$$\alpha_{t+1}(j) = \left[\sum_{i=1}^{N} \alpha_t(i) a_{ij}\right] b_j(o_{t+1})$$

III. 终止 $$P(O|\lambda) = \sum_{i=1}^{N} \alpha_T(i)$$

后向算法

后向变量 $\beta_t(i) = P(O_{t+1}, O_{t+2}, \ldots, O_T, q_t=\theta_i|\lambda)$ $1 \le t \le T-1$

算法步骤

I. 初始化： $\beta_T(i) = 1$

II. 迭代： $\beta_t(i) = \sum_{j=1}^{N} a_{ij} b_j(o_{t+1}) \beta_{t+1}(j)$

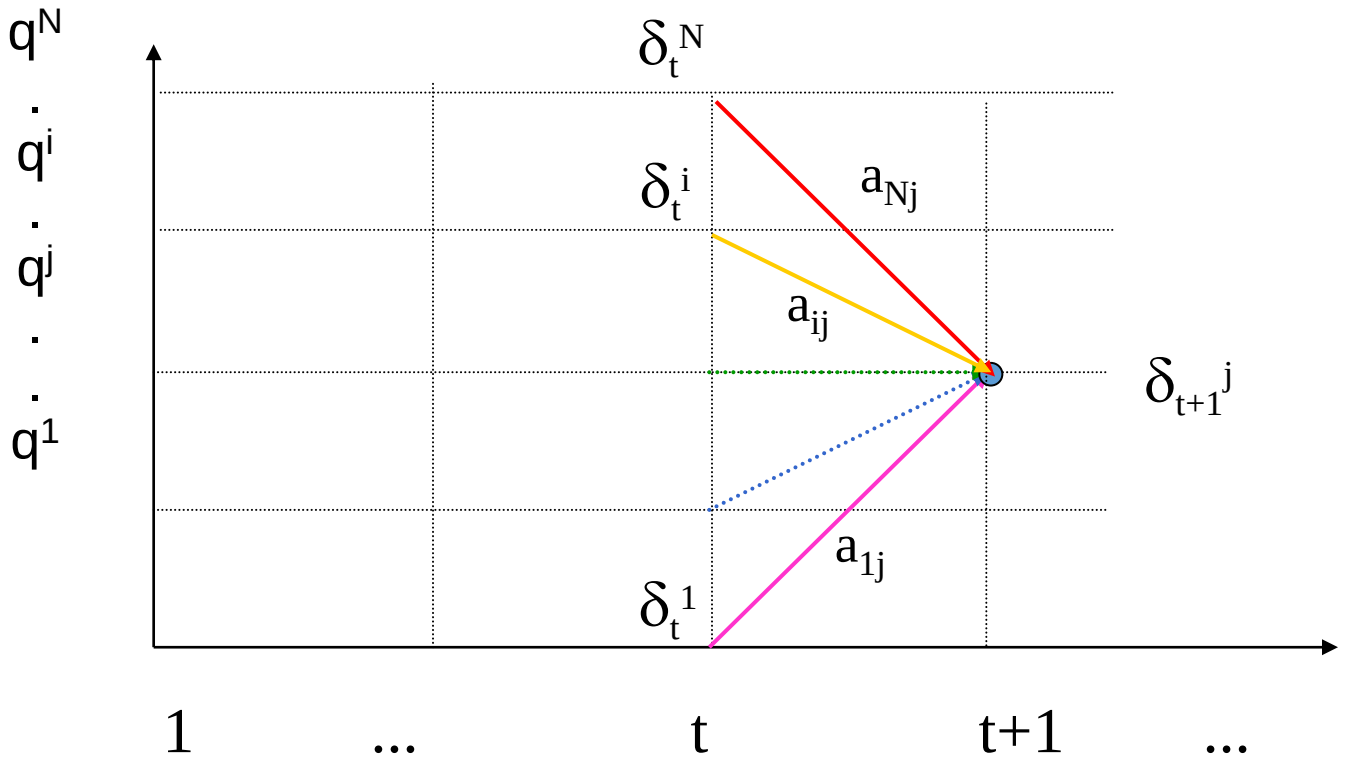
III. 终止 $P(O|\lambda) = \sum_{i=1}^{N} \beta_1(i)$

隐马尔可夫模型的三个基本问题（2）

- Viterbi Algorithm（解码）
- 给定训练好了的隐马尔可夫模型参数λ
- 给定该模型的一个观测序列o
- 求解：生成该序列的内部状态S = q1, q2, …qT
- S是能够最为合理的解释观测序列o的状态序列
- 用于分析HMM中状态的具体含义以及序列分割等

$$\delta_k(i) = \max_{q_1,q_2,\dots,q_{t-1}} P[q_1, q_2, \dots, q_{t-1}, q_t = i, O_1, O_2, \dots, O_t|\lambda]$$

$$\delta_t(j) = \max_{1 \le i \le N} \left[\delta_{t-1}(i) a_{ij} \right] b_j(O_i), \qquad 2 \le t \le T, 1 \le j \le N$$



Viterbi Algorithm

- 初始化：

$$\delta_1(i) = \pi_i b_i(O_1), \quad 1 \le \mathrm{i} \le \mathrm{N}$$

$$\varphi_1(i) = 0, \quad 1 \le \mathrm{i} \le \mathrm{N}$$

- 递归：

$$\delta_t(j) = \max_{1 \le i \le N}[\delta_{t-1}(i)a_{ij}]b_j(O_i), \quad 2 \le t \le T, 1 \le j \le N$$

- 终结：

$$\varphi_t(j) = \arg\max_{1 \le i \le N}[\delta_{t-1}(i)a_{ij}], \quad 2 \le t \le T, 1 \le j \le N$$

- 求S序列：

$$P^* = \max_{1 \le i \le N}[\delta_T(i)])$$

$$q_T^* = \arg\max_{1 \le i \le N}[\delta_T(i)]$$

$$q_t^* = \varphi_{t+1}(q_{t+1}^*), \quad t = T-1, T-2, \ldots, 1$$

隐马尔可夫模型的三个基本问题（3）

- 学习问题：模型训练
- 给定训练样本的集合O，训练隐马尔可夫模型参数λ，使得$P(O|\lambda)$最大
- Baum-Welch Algorithm

L. R. Rabiner, A tutorial on hidden Markov models and selected applications in speech recognition, Proceedings of the IEEE, vol. 77, no. 2, pp. 257-286,1989.

Baum-Welch Algorithm

- 基本思路与步骤：

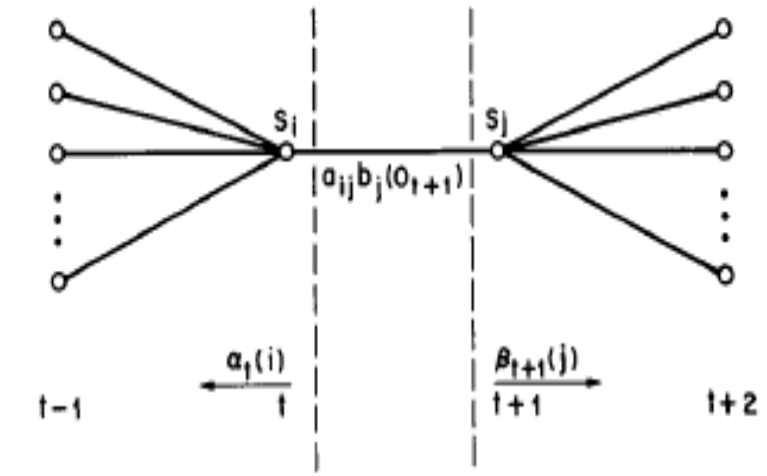
1. 初始模型（待训练模型）λ_0,
2. 基于λ_0以及观察值序列O，训练新模型λ；
3. 如果 $\log(P(O|\lambda)) - \log(P(O|\lambda_0) < Delta$，说明训练已经达到预期效果， 算法结束。
4. 否则，令$\lambda_0 = \lambda$ ，继续第2步工作

Baum-Welch Algorithm

给定模型λ，观测序列条件下

从i到j的转移概率$\xi_t(i,j)$定义为：

$$P(it = q_i, i_{t+1} = q_j, O|\lambda) = \alpha_t(i)a_{ij}b_j(o_{t+1})\beta_{t+1}(j)$$



$$\xi_t(i,j) = \frac{P(it = q_i, i_{t+1} = q_j, O|\lambda)}{P(O|\lambda)}$$

$$= \frac{\alpha_t(i)a_{ij}b_j(o_{t+1})\beta_{t+1}(j)}{\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_t(i)a_{ij}b_j(o_{t+1})\beta_{t+1}(j)}$$

其中，$$\gamma_t(i) = P(i_t = q_i|O,\lambda) = \frac{\alpha_t(i)\beta_t(i)}{\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_t(i)a_{ij}b_j(o_{t+1})\beta_{t+1}(j)}$$

$$a_{ij} = \frac{\sum_{t=1}^{T-1}\xi_t(i,j)}{\sum_{t=1}^{T-1}\gamma_t(i)} \qquad b_j(k) = \frac{\sum_{t=1,o_t=v_k}^{T-1}\gamma_t(j)}{\sum_{t=1}^{T}\gamma_t(j)} \qquad \pi_i = \gamma_1(i)$$

基于深度学习的行为识别

Karen Simonyan and Andrew Zisserman , Two-Stream Convolutional Networks for Action Recognition in Videos, NIPS 2014

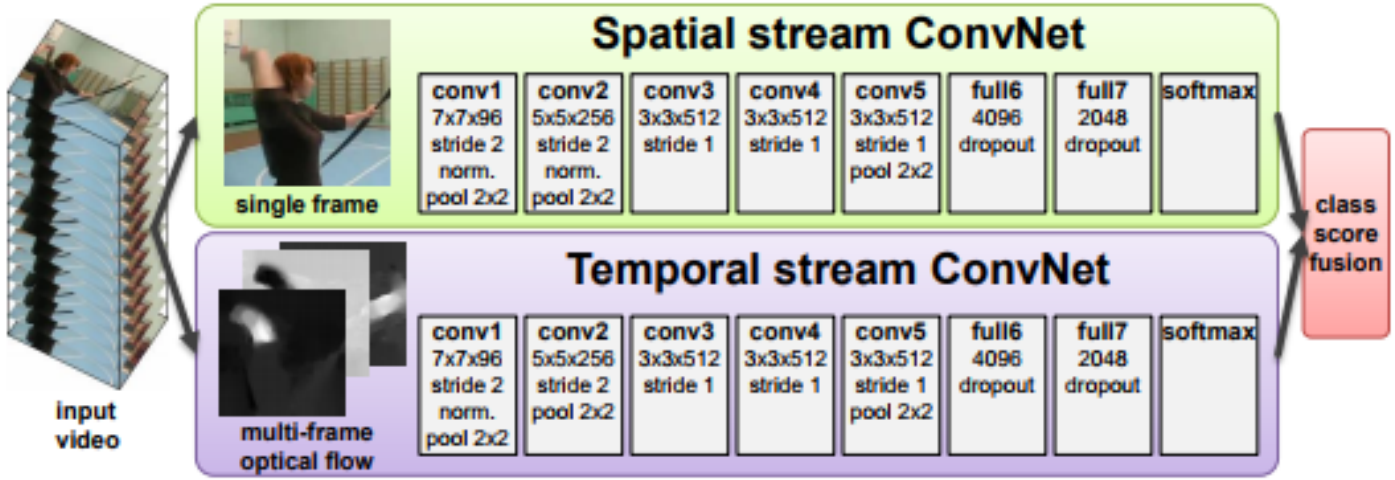


Figure 1: **Two-stream architecture for video classification.**

Spatial stream ConvNet operates on individual video frames, effectively performing action recognition from still images. The static appearance by itself is a useful clue, since some actions are strongly associated with particular objects. In fact, as will be shown in Sect. 6, action classification from still frames (the spatial recognition stream) is fairly competitive on its own. Since a spatial ConvNet is essentially an image classification architecture, we can build upon the recent advances in large-scale image recognition methods [15], and pre-train the network on a large image classification dataset, such as the ImageNet challenge dataset. The details are presented in Sect. 5. Next, we describe the temporal stream ConvNet, which exploits motion and significantly improves accuracy.

基于深度学习的行为识别

Shuiwang Ji, Wei Xu, Ming Yang and Kai Yu, 3D convolutional neural networks for human action recognition, PAMI 2013

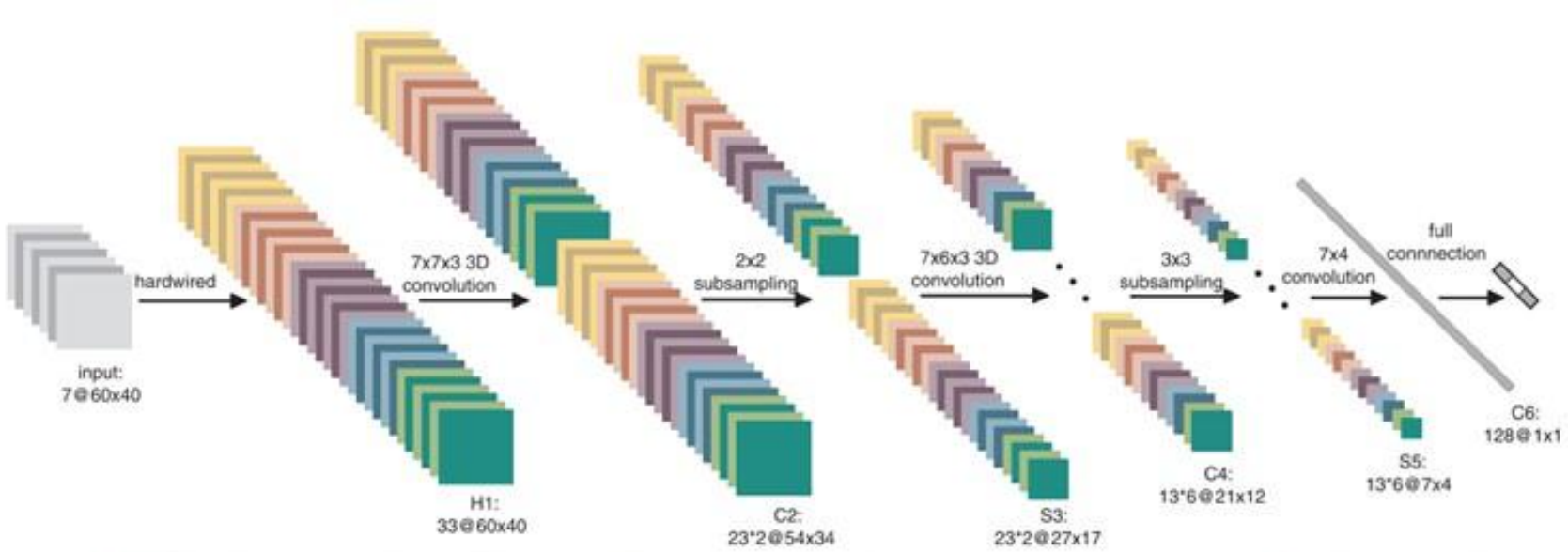
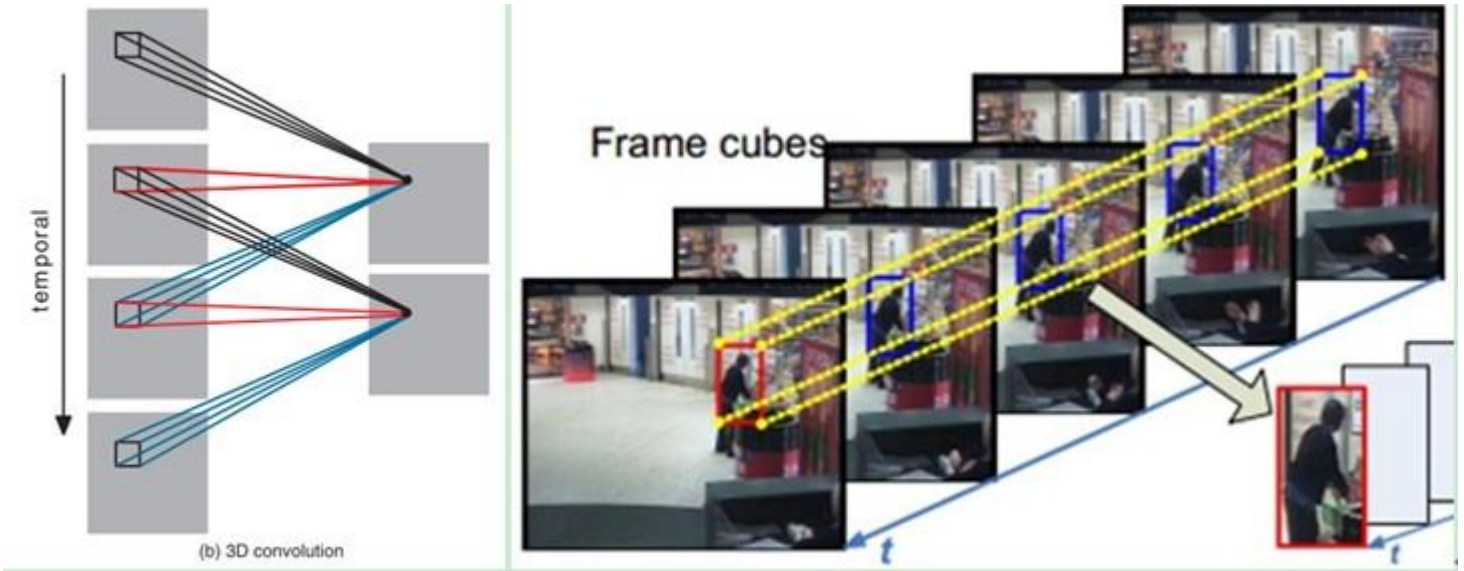


Fig. 3. A 3D CNN architecture for human action recognition. This architecture consists of one hardwired layer, three convolution layers, two subsampling layers, and one full connection layer. Detailed descriptions are given in the text.

基于深度学习的行为识别

Shuiwang Ji, Wei Xu, Ming Yang and Kai Yu, 3D convolutional neural networks for human action recognition, PAMI 2013



1 背景内容

2 运动表达

3 行为识别

4 小节

小节

- 运动表达
- 基于DNN的物体表达
- 稀疏、低秩表达
- 行为识别
  - 常规方法
  - 基于DNN的方法

课后练习

1试编程实现基于稀疏表达的人脸识别。

2试编程实现DeepID。

谢谢！