

Примеры практического использования алгоритмов минимизации энергии

Спецкурс «Структурные методы анализа изображений и
сигналов»

Марковские случайные поля

$$E(X) = \sum_{p \in \mathcal{P}} D_p(x_p) + \sum_{(p,q) \in \mathcal{E}} V_{pq}(x_p, x_q), \quad x_p = 1, \dots, K$$
$$\min_X E(X) = ???$$

- В общем случае задача NP-трудная
- Задача полиномиально разрешима в частных случаях:
 - Граф \mathcal{E} – дерево (динамическое программирование)
 - $K = 2$ и функция субмодулярна (разрезы графов)
- Приближенные алгоритмы минимизации энергии:
 - Tree-ReWeighted Message Passing (TRW)
 - α -расширение

План лекции

- 1 Восстановление изображений
- 2 Сегментация изображений
- 3 Стерео
- 4 Панорамы
- 5 Поиск составных объектов на изображении

Восстановление изображений



$$E(X) = \sum_{p \in \mathcal{P}} D_p(x_p, I_p) + \alpha \sum_{(p,q) \in \mathcal{E}} V_{pq}(x_p, x_q), \quad x_p = 0, \dots, 255$$

x_p – цвета очищенного изображения

Восстановление изображений: унарные потенциалы

Унарные потенциалы связаны с моделью шума на изображении:

Унарные потенциалы	Распределение шума
$[x_p \neq I_p]$ $(x_p - I_p)^2$ $ x_p - I_p $ $\min((x_p - I_p)^2, c)$ $\min(x_p - I_p , c)$	равномерное нормальное Лапласа

I_p – цвета зашумленного изображения

Восстановление изображений: бинарные потенциалы

Бинарные потенциалы $V_{pq}(x_p, x_q)$:

- Выпуклый штраф: $|x_p - x_q|$ или $(x_p - x_q)^2$
- Модель Поттса: $[x_p \neq x_q]$
- Усеченный выпуклый штраф: $\min((x_p - x_q)^2, c)$ или $\min(|x_p - x_q|, c)$

Сегментация изображений



$$E(X) = \sum_{p \in \mathcal{P}} D_p(x_p) + \alpha \sum_{(p,q) \in \mathcal{E}} V_{pq}(x_p, x_q), \quad x_p = 1, \dots, K$$

x_p – метки классов пикселей

Сегментация изображений: унарные потенциалы

$$D_p(x_p) = -\log P(y_p | x_p).$$

U_p – вектор всех локальных характеристик изображения:

- цвет
- позиция
- текстурные признаки

Сегментация изображений: бинарные потенциалы

Бинарные потенциалы – регуляризатор, отвечающий за “гладкость” решения. Обычно потенциалы $V_{pq}(x_p, x_q)$ выбирают так:

- Модель Поттса: $[x_p \neq x_q]$. Приводит к наиболее коротким границам.
- Обобщенная модель Поттса: $c_{pq} [x_p \neq x_q]$, где $c_{pq} = A + B \exp\left(-\frac{\|I_p - I_q\|^2}{2\sigma^2}\right)$. Границы проходят там, где есть скачки цвета.

Откуда брать модели объектов?

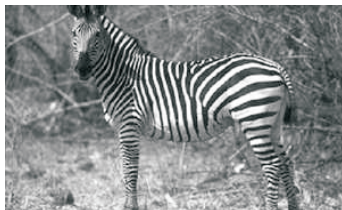
Постановки задачи сегментации:

- Сегментация с учителем.
- Сегментация без учителя.
- Интерактивная сегментация.

Сегментация с учителем

- Метки классов соответствуют объектам с определенным внешним видом.
- Есть размеченная обучающая выборка.
- Унарные и бинарные потенциалы настраиваются по выборке.
- В унарных потенциалах можно использовать любой классификатор, дающий вероятностные выходы.
- Бинарные потенциалы – кросс-валидация или структурное обучение.

Сегментация без учителя

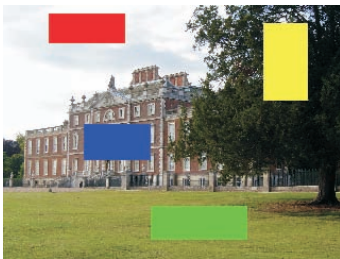


- 1 Выбрать случайные патчи и сгенерировать модели объектов.
- 2 Сегментировать.
- 3 Пересчитать модели.
- 4 Повторять 2 и 3 до сходимости.

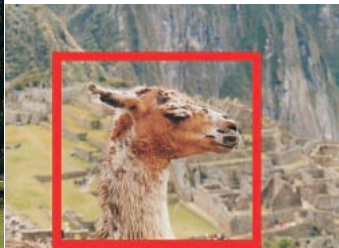
Интерактивная сегментация

Пользователь предоставляет дополнительную информацию об объектах на изображении.

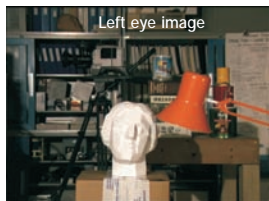
“Семена” объектов



Рамка (GrabCut)



Стерео



$$E(X) = \sum_{p \in \mathcal{P}} D_p(x_p) + \alpha \sum_{(p,q) \in \mathcal{E}} V_{pq}(x_p, x_q), \quad x_p = -D, \dots, 0, \dots, D$$

x_p – сдвиг пикселя до соответствия (диспаритет)

Стерео: унарные потенциалы

Унарные потенциалы показывают насколько соответствующие пиксели хорошо подходят друг к другу.

Разность квадратов интенсивностей: $D_p(x_p) = (I_p - I'_{p+x_p})^2$.

Работает плохо из-за ошибок дискретизации.

Более устойчивые величины:

$$C_{fwd}(p, d) = \min_{d-0.5 \leq x \leq d+0.5} |I_p - I'_{p+x}|,$$

$$C_{rev}(p, d) = \min_{p-0.5 \leq x \leq p+0.5} |I_x - I'_{p+d}|.$$

Итоговый потенциал:

$$D_p(x_p) = (\min(C_{fwd}(p, x_p), C_{rev}(p, x_p), const))^2$$

Стерео: бинарные потенциалы

Обобщенные потенциалы Поттса:

$$V_{pq}(x_p, x_q) = c_{p,q}[x_p \neq x_q], p, q \geq 0.$$

$$c_{p,q} = \begin{cases} C_1, & |I_p - I_q| < T, \\ C_2, & \text{иначе.} \end{cases}$$

Обычно T – небольшое, $C_1 > C_2$.

Панорамы



$$E(X) = \sum_{p \in \mathcal{P}} D_p(x_p) + \alpha \sum_{(p,q) \in \mathcal{E}} V_{pq}(x_p, x_q), \quad x_p = 1, \dots, K$$

x_p – номер изображения, из которого берется пиксель p

Панорамы: потенциалы

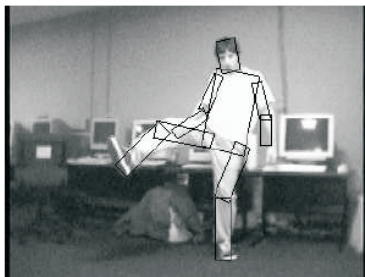
Унарные потенциалы не 0 только для семян:

$$D_p(x_p) = \begin{cases} +\infty, & p \in \text{Seed}_k, \quad k \neq x_p, \\ 0, & \text{иначе.} \end{cases}$$

Бинарные потенциалы:

- $V_{pq}(x_p, x_q) = |I_p(x_p) - I_p(x_q)| + |I_q(x_p) - I_q(x_q)|,$
- $V_{pq}(x_p, x_q) = \frac{|I_p(x_p) - I_p(x_q)| + |I_q(x_p) - I_q(x_q)|}{|\nabla_{pq} I(x_p)| + |\nabla_{pq} I(x_q)|}$

Поиск составных объектов на изображении



$$E(X) = \sum_{p \in \mathcal{P}} D_p(x_p) + \alpha \sum_{(p,q) \in \mathcal{E}} V_{pq}(x_p, x_q), \quad x_p = 1, \dots, K$$

x_p – позиция части номер p на изображении

Поиск объектов: потенциалы

Модель лица:



Параметры каждой части – позиция.

Унарные потенциалы:

$$D_p(\mathbf{x}_p) = -\log \mathcal{N}(\beta(\mathbf{x}_p) | \mu_p, \Sigma_p)$$

$\beta(\mathbf{x}_p)$ – линейные фильтры, примененные в точке \mathbf{x}_p

Бинарные потенциалы:

$$V_{pq}(\mathbf{x}_p, \mathbf{x}_q) = -\log \mathcal{N}(\mathbf{x}_p - \mathbf{x}_q | \mu_{pq}, \Sigma_{pq})$$

Прочие задачи

- Сегментация движений
- Синтаксический разбор предложений
- Трехмерная реконструкция
- Восстановление текстуры
- ...