Ластовичек Дмитрий. АД
БММО
Практическое задание №2

1 Теория

E-шаг

$$p(d_k|X_k, \theta, A) = \frac{p(X_k|d_k, \theta)p(d_k|A)}{\sum_{d_k} p(X_k|d_k, \theta)p(d_k|A)}$$

M-шаг

$$\begin{aligned} &E_q \log P(X, d|\theta, A) = E_q \sum_k [-\frac{1}{2s^2} \sum_{i,j \in face} (X_{ij}^k - F_{i-di,j-dj})^2 \\ &-\frac{1}{2s^2} \sum_{i,j \notin face} (X_{ij}^k - B_{i,j})^2 + HW \log \frac{1}{\sqrt{2\pi s^2}} + \log A(d_i^k, d_j^k)] \end{aligned} \tag{1}$$

A:
Входит только в последнее слагаемое, поэтому

$$\begin{cases} \sum_k \sum_{i,j} \log A_{ij} q_{ij}^k \to max_A \\ \sum_{i,j} A_{ij} = 1 \end{cases} \tag{2}$$

Решаем методом Лагранжа:

$$L = \sum_k \sum_{i,j} \log A_{ij} q_{ij}^k + \lambda(\sum_{i,j} A_{ij} - 1)$$

$$L'_{A_{ij}} = \frac{\sum_k q_{ij}^k}{A_{ij}} + \lambda = 0$$

$$A_{ij} = -\frac{\sum_k q_{ij}^k}{\lambda}$$

$$\sum_{i,j} A_{ij} = -\frac{\sum_{i,j} \sum_k q_{ij}^k}{\lambda} = 1$$

$$A_{ij} = -\frac{\sum_k q_{ij}^k}{\sum_{i,j} \sum_k q_{ij}^k}$$

F:
Входит только в первое слагаемое

$$\sum_k E_q(d_k)[-\frac{1}{2s^2} \sum_{i,j \in face} (X_{ij}^k - F_{i-d_i^k, j-d_j^k})^2] \to max_F$$

$$\sum_k \sum_{d_i,d_j} q_{d_i,d_j}^k \sum_{i,j \in face} (X_{ij}^k - F_{i-di,j-dj})^2 \to min_F$$

$$\sum_{d_i,d_j} \sum_k q_{d_i,d_j}^k \sum_{i=0}^{h-1} \sum_{j=0}^{w} (X_{i+d_i,j+d_j}^k - F_{i,j})^2 \to min_F$$

$$\sum_{i=0}^{h-1} \sum_{j=0}^{w} \sum_{d_i,d_j} \sum_k q_{d_i,d_j}^k (X_{i+d_i,j+d_j}^k - F_{i,j})^2 \to min_F$$

Приравнивая производную к нулю по каждому F_{ij} к нулю получим

$$F_{i,j} = \frac{\sum_{d_i,d_j} \sum_k q_{d_i,d_j}^k X_{i+d_i,j+d_j}^k}{\sum_{d_i,d_j} \sum_k q_{d_i,d_j}^k}$$

B:
Входит только во второе слагаемое:

$$\sum_k E_q(d_k)[-\frac{1}{2s^2} \sum_{i,j \notin face} (X_{ij}^k - B_{i,j})^2] \to max_F$$

$$\sum_k \sum_{d_i,d_j} q_{d_i,d_j}^k \sum_{i,j \notin face} (X_{ij}^k - B_{ij})^2 \to min_F$$

$$\sum_{i=0}^{H-1} \sum_{j=0}^{W-1} \sum_{d_i,d_j} \sum_k q_{d_i,d_j}^k (X_{ij}^k - F_{i,j})^2 I(i,j \notin face) \to min_F$$

Приравнивая производную к нулю по каждому B_{ij} к нулю получим

$$B_{i,j} = \frac{\sum_{d_i,d_j} \sum_k q^k_{d_i,d_j} X^k_{i,j} I(i,j \notin face)}{\sum_{d_i,d_j} \sum_k q^k_{d_i,d_j} I(i,j \notin face)}$$

s:
Входит в три слагаемых:

$$R^k = \sum_{i=0}^{h-1} \sum_{j=0}^{w} \sum_{d_i,d_j} \sum_k q^k_{d_i,d_j} (X^k_{i+d_i,j+d_j} - F_{i,j})^2$$

$$+ \sum_{i=0}^{H-1} \sum_{j=0}^{W-1} \sum_{d_i,d_j} \sum_k q^k_{d_i,d_j} (X^k_{ij} - F_{i,j})^2 I(i,j \notin face)$$

$$\sum_k -\frac{R^k}{2s^2} - HWK \log s \rightarrow max_s$$

Приравниваем к нулю производную,получим

$$s = \sqrt{\frac{\sum_k R^k}{HWK}}$$

Для MAP-EM просто нужно преобразовать формулы, учитывая, что $q_{d^k_{ij}}$ равна 1 в одной точке

Нижняя оценка L

$$L = \sum_k \sum_{d_i,d_j} [\log p(X_k|d_i,d_j,\theta) + logA_{d_i,d_j} - logq^k_{d_i,d_j}] q^k_{d_i,d_j}$$

2 Анализ

2.1 Сгенерированые данные

Сгенерируем тестовые F_{test}, B_{test}. Позиция смайлика выбирается равномерно на фоне для каждого образца. Параметр шума задаем

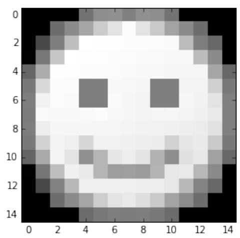

Рис. 2.1.1: F_{test}

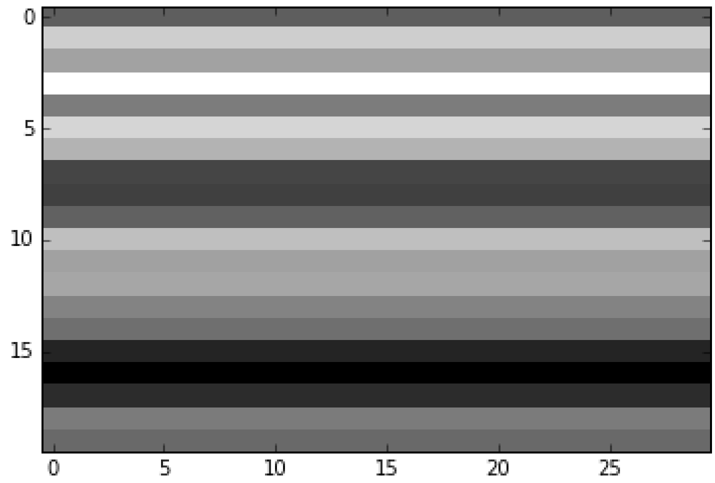

Рис. 2.1.2: B_{test}

Сильно ли влияет начальное приближение на параметры на результаты работы?
Начальное приближение влияет на скорость сходимости и на саму сходимость алгоритма. При плохих начальных приближениях знаменательв формуле пересчета для B может обратиться в нуль, и метод возратит NaN. Поэтому в данной работе я использовал начальное приближение основанное на данных X (генерируем начальные F и B как нормальное распределение с в каждой точке i, j как $avg(X_{ij}), std(X_{ij})$

Стоит ли для данной задачи запускать EM алгоритм из разных начальных приближений?
При выбранных мной начальных приближениях на F и B, описанных выше, алгоритм сходился почти всегда (За исключение, когда в знаменателе для B 0). Поэтому запускать из разных приближений стоит.

2.2 Тестирование на сгенерированных данных

s=10
В данном случае, т.к. s маленький быстро удается определить точное положение смайлика и зануляется знаменатель в B. Поэтому приходится делать несколько запусков.

Число образцов = 500. L=-1118552.85311.

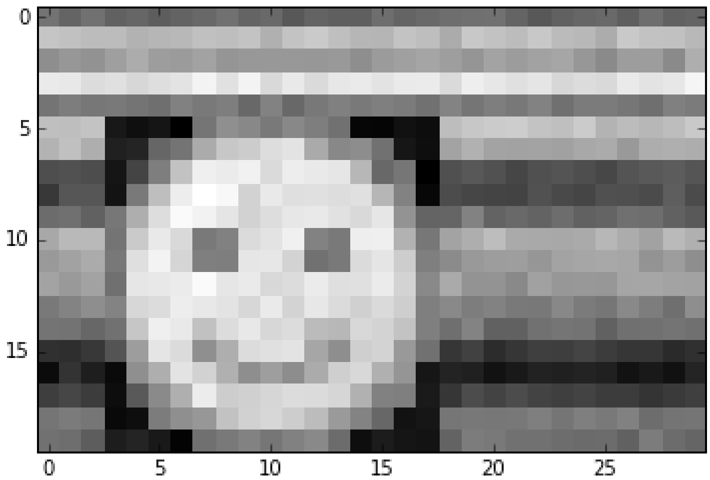

Рис. 2.2.1: $X_{test} example$

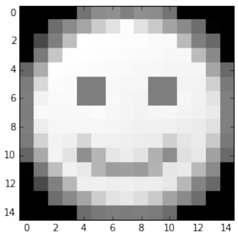

Рис. 2.2.2: $F_{test} 500 samples$

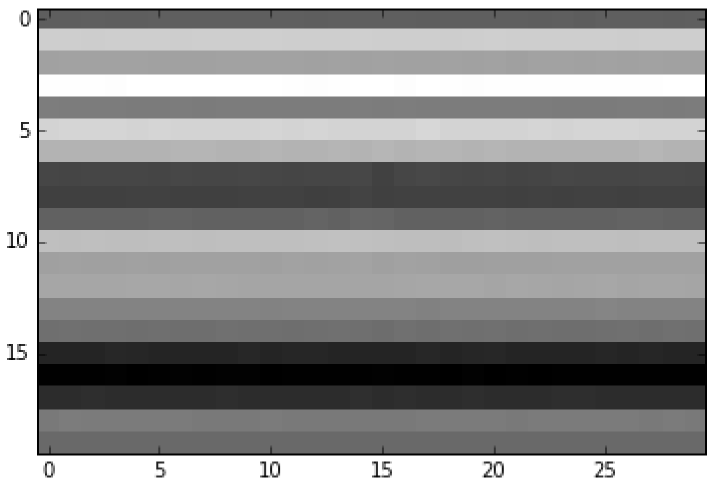

Рис. 2.2.3: $B_{test} 500 samples$

Число образцов = 100. В данном случае образцов мало, и может быть ситуация что смайлик всегда перекрывает середину изображения, поэтому фон в центре невозможно определить и запуски часто не сходятся, а сходятся лишь для не совсем верного F (который "якобы"не перекрывает центр).

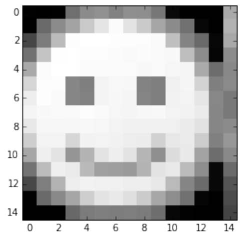

Рис. 2.2.4: $F_{test} 500 samples$

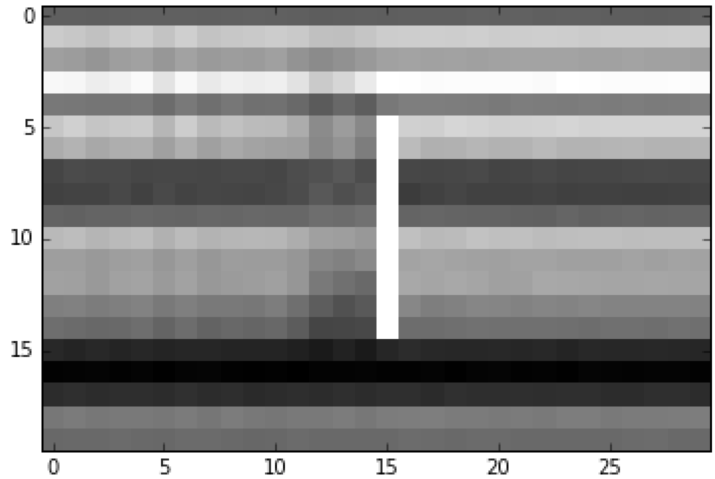

Рис. 2.2.5: $B_{test} 500 samples$

Далее для фиксированного числа образцов (500) буду увеличивать шум

s=50. L=-1601106.65413

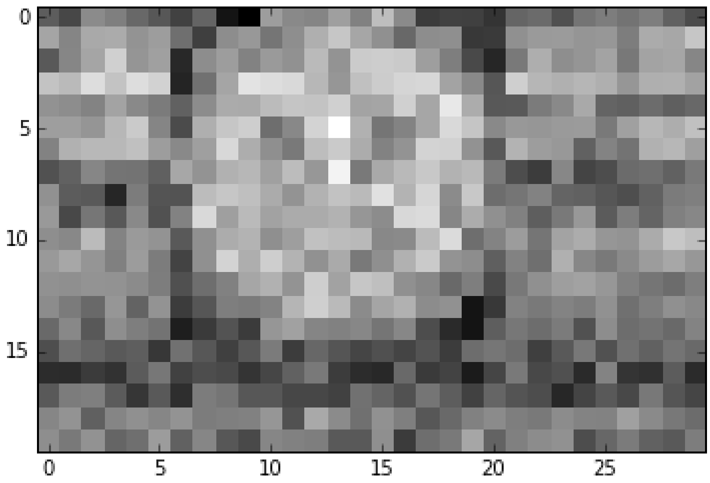

Рис. 2.2.6: $X_{test} example$

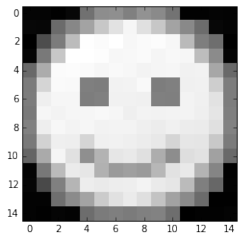

Рис. 2.2.7: F_{test}

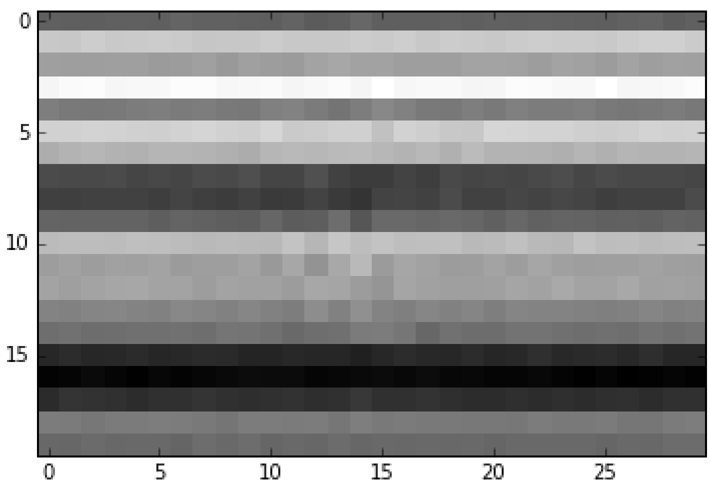

Рис. 2.2.8: B_{test}

s=100. L=-1809391.78374

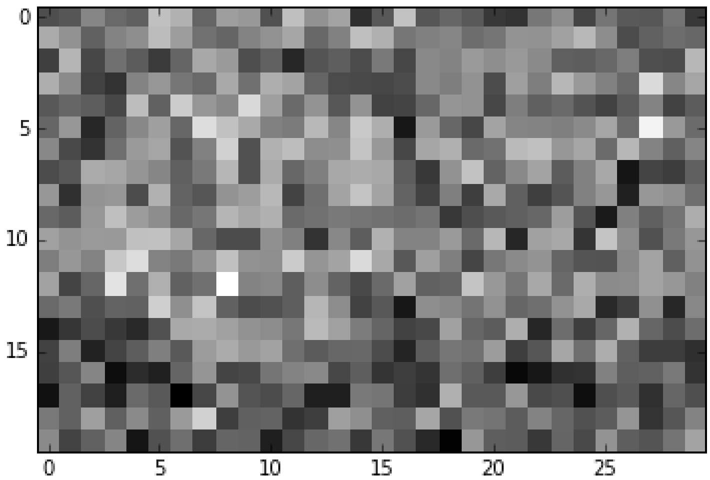

Рис. 2.2.9: $X_{test} example$

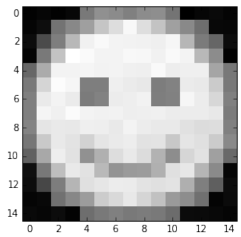

Рис. 2.2.10: F_{test}

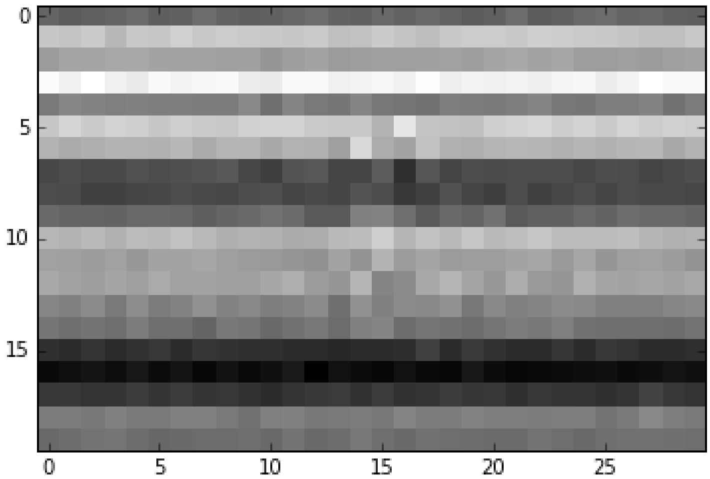

Рис. 2.2.11: B_{test}

s=150. L=-1930956.92011

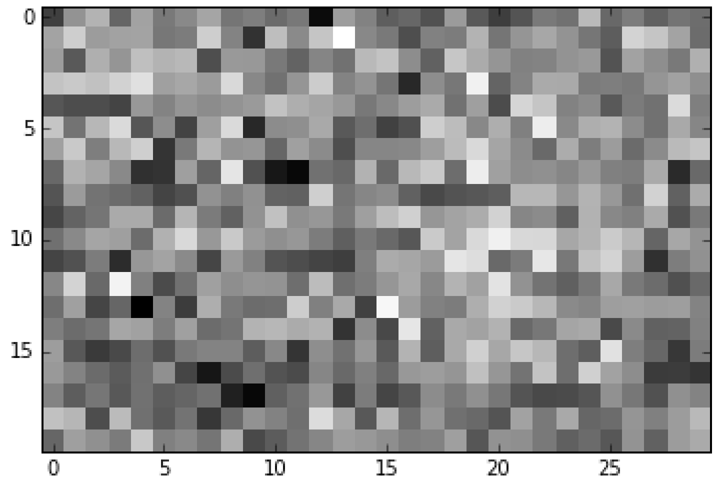

Рис. 2.2.12: $X_{test} example$

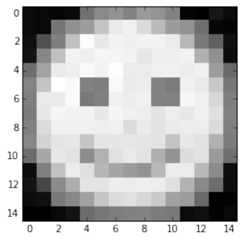

Рис. 2.2.13: F_{test}

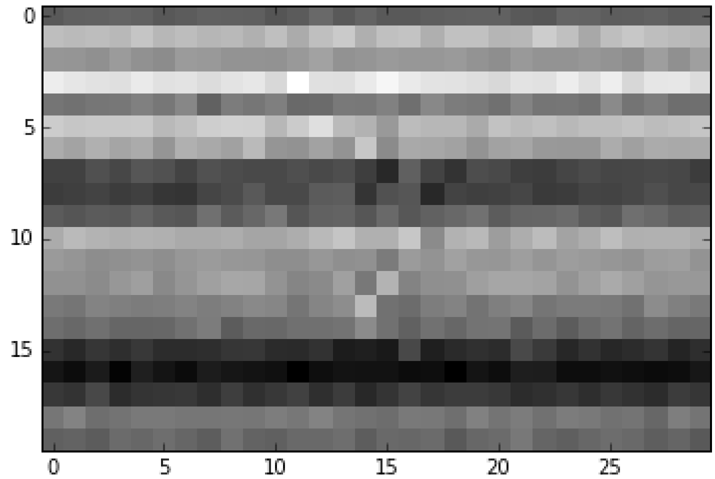

Рис. 2.2.14: B_{test}

s=200. L=-2016634.89901

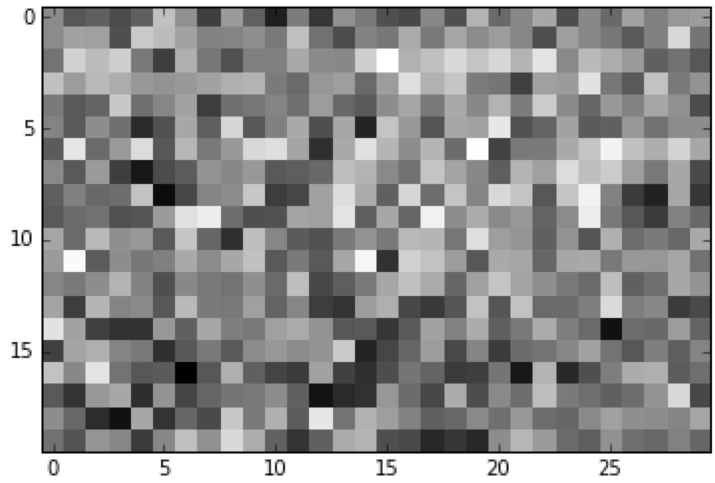

Рис. 2.2.15: $X_{test} example$

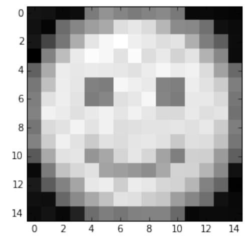

Рис. 2.2.16: F_{test}

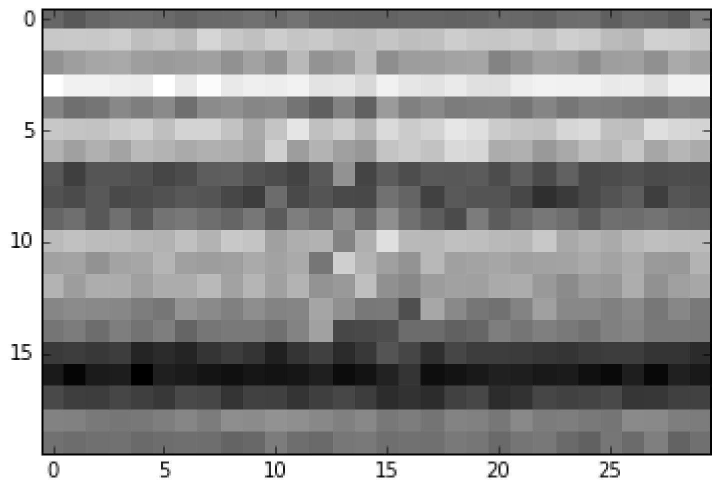

Рис. 2.2.17: B_{test}

s=300. L=-2138216.35902

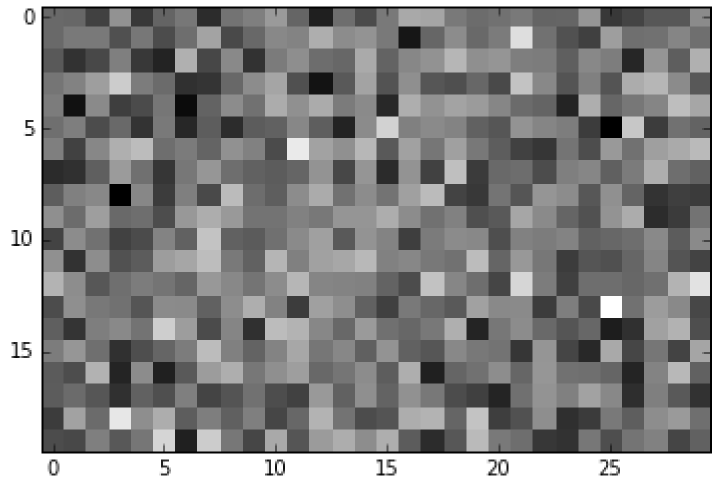

Рис. 2.2.18: $X_{test} example$

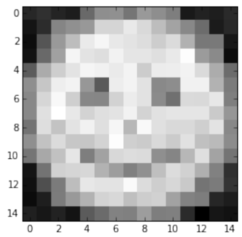

Рис. 2.2.19: F_{test}

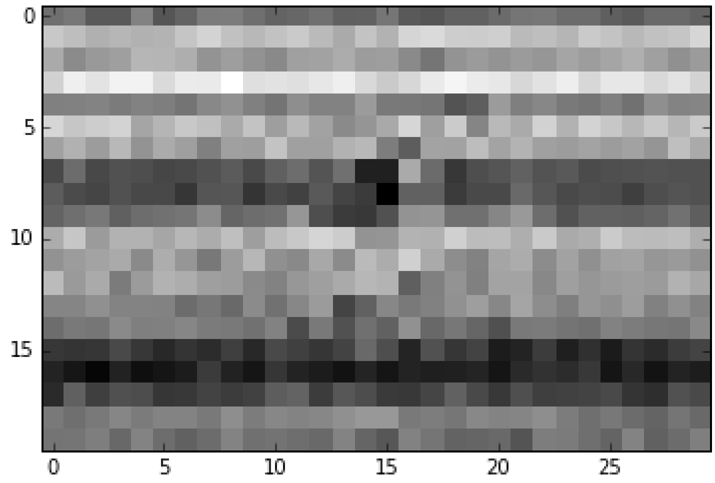

Рис. 2.2.20: B_{test}

s=400. L=-2224396.74142

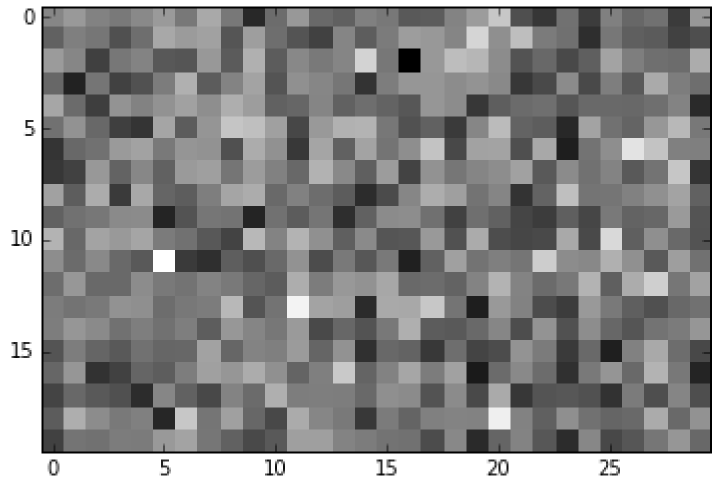

Рис. 2.2.21: $X_{test} example$

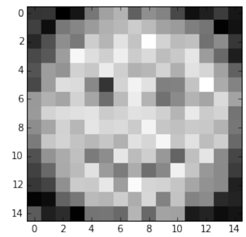

Рис. 2.2.22: F_{test}

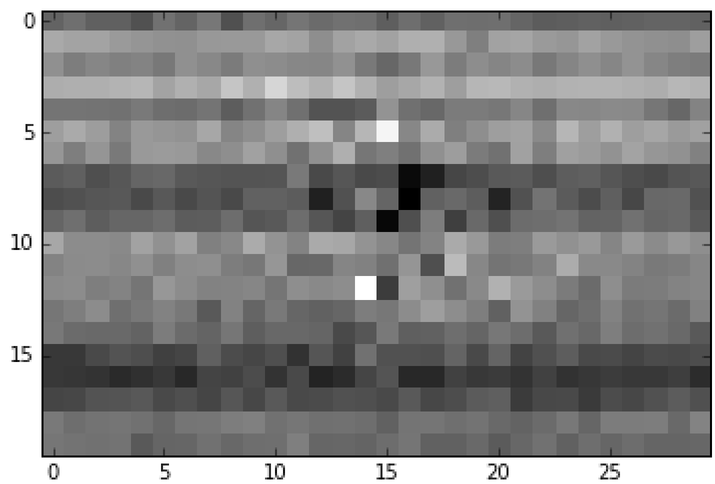

Рис. 2.2.23: B_{test}

s=600. L=-2224396.74142

Здесь уже объекты не различимы. Заметим что правдоподобие все время уменьшалось с увеличением s, что очевидно.

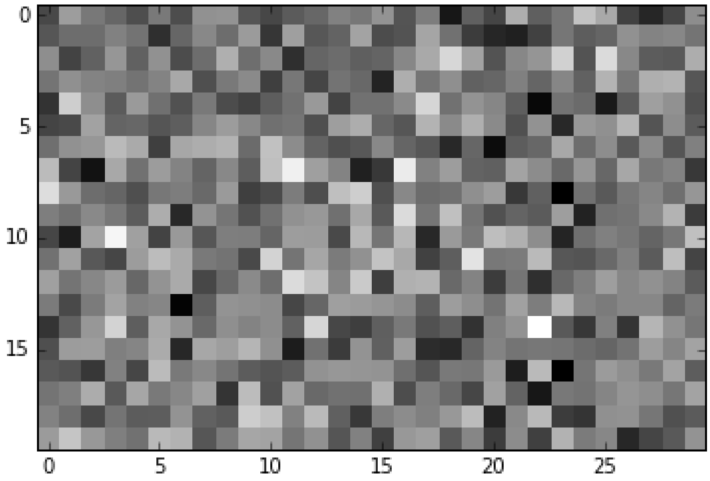

Рис. 2.2.24: $X_{test}example$

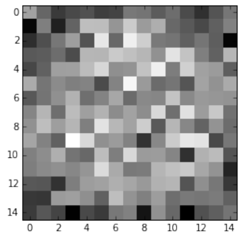

Рис. 2.2.25: F_{test}

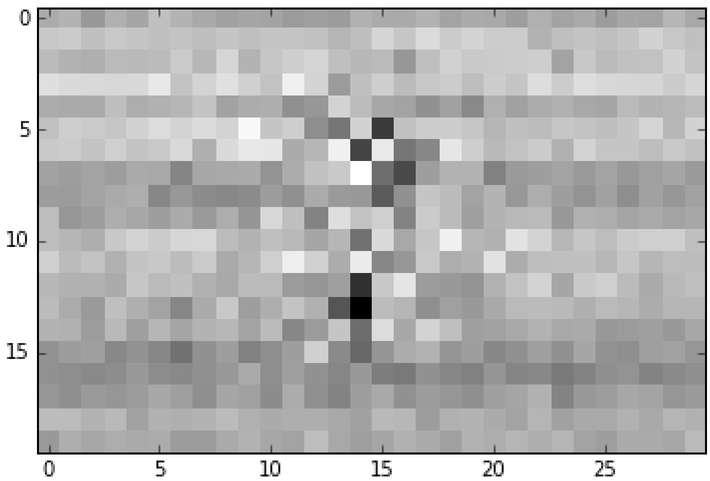

Рис. 2.2.26: B_{test}

2.3 EM vs MAP-EM

$L_{MAP-EM} = -807325.715869$
$L_{EM} = -806741.622603$ Получаем, что MAP-EM работает несколько хуже, чо очевидно, учитывая, что мы сделали приближенное упрощение модели.

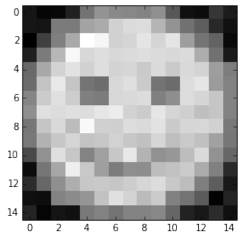

Рис. 2.3.1: f_{em}

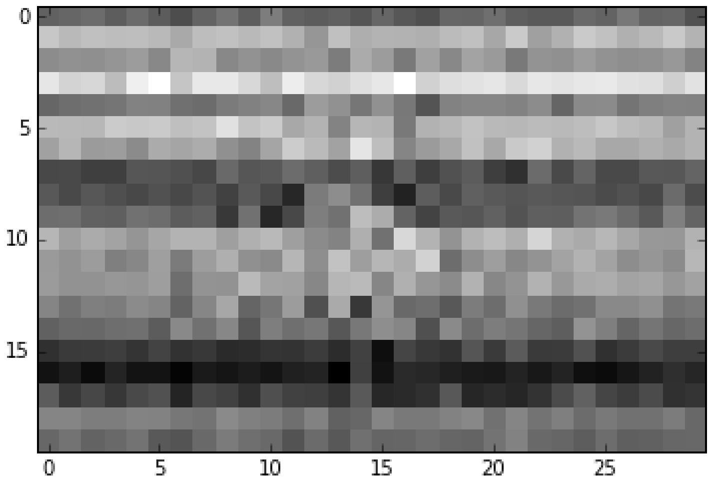

Рис. 2.3.2: b_{em}

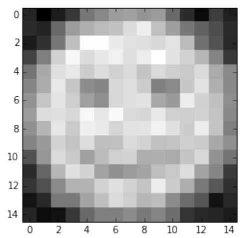

Рис. 2.3.3: f_{map-em}

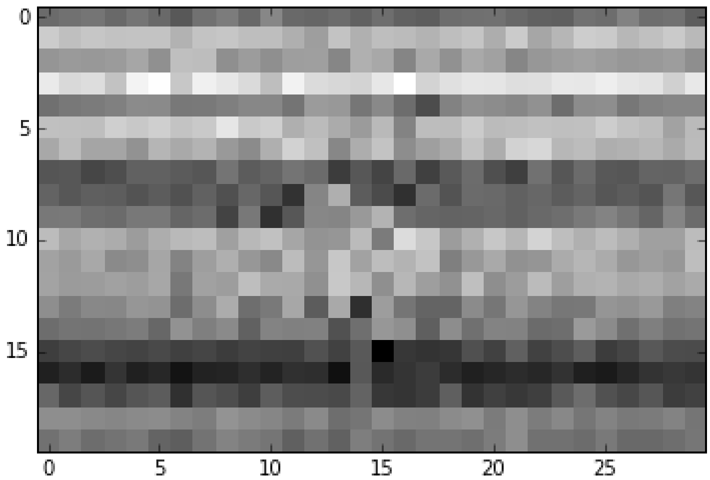

Рис. 2.3.4: b_{map-em}

2.4 Преступник

На всей выборке получаем:

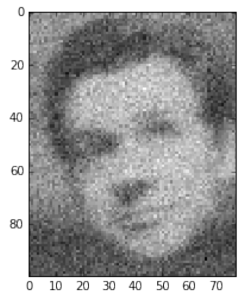

Рис. 2.4.1: F

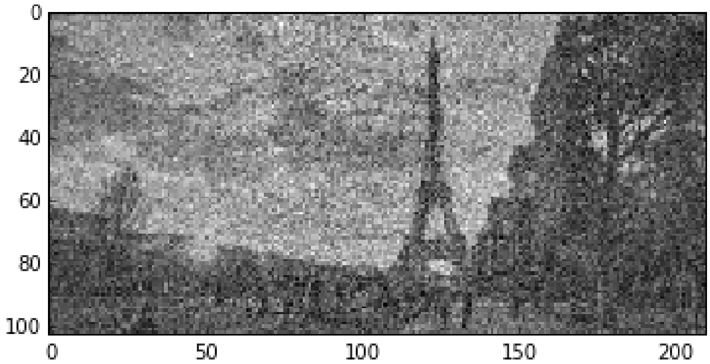

Рис. 2.4.2: B

На выборке 200 образцов получаем:

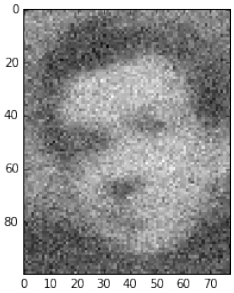

Рис. 2.4.3: F

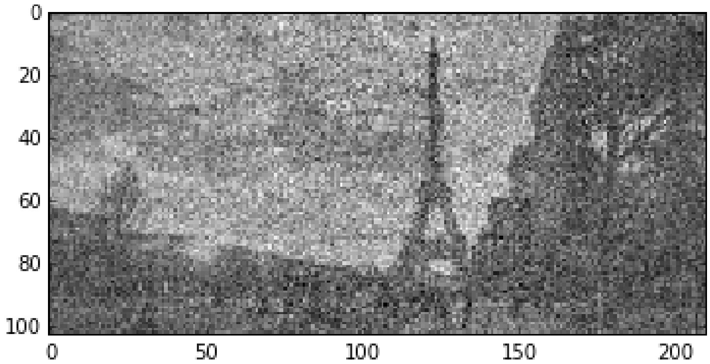

Рис. 2.4.4: B

2.5 Модификации

Единственное, что я сделал для ускорения программы - немного оптимизировал вычисление lpx_d и s (изначально вычисление lpx_d и s занимало порядка 350 секунд на 1 итерации, после оптимизации кода - 130 секунд) Суть оптимизации заключается в том, что мы считаем отклонение X от B (в области вне $face$) не каждый раз, а пересчитываем по сравнению с предыдущим шагом - например dj сместилось вправо и мы пересчитываем отклонение лишь по изменению покрываемой области $face$ на данном шаге.
Еще можно распаралелить все циклы по образцам и позиции начала престуника.