

Модификации одномерных скрытых марковских моделей для задачи распознавания лиц

Т. А. Гульятеева, А. А. Попов

Новосибирский государственный технический университет, Новосибирск, Россия

gult@mail.ru, alex@fpm.ami.nstu.ru

Аннотация

В данной работе представлен один из статистических методов распознавания изображений лиц людей, основанный на одномерных скрытых марковских моделях. Рассмотрены некоторые модификации моделей, увеличивающие распознающие свойства системы.

Ключевые слова: скрытые марковские модели, распознавание лиц.

1. ВВЕДЕНИЕ

В настоящее время наблюдается непреходящий интерес к проблеме распознавания лиц. Под распознаванием будем понимать идентификацию изображения неизвестного лица с одной из известных персон. Данная задача актуальна как в области интеллектуальных сред, так и в системах безопасности. Одни из первых разработанных методов распознавания лиц – метод главных компонент (собственных лиц), отличительной особенностью которого является то, что главные компоненты несут в себе информацию о признаках некоторого обобщенного лица. Распознавание лиц с использованием линейного дискриминантного анализа основывается на предположении о линейной разделимости классов (персон) в пространстве изображений. Нейросетевые методы обладают хорошей обобщающей способностью. Обзор этих и других методов распознавания (таких как сравнение эластичных графов, сравнение эталонов и др.) приведен в [1]. Далее мы рассмотрим распознавание лиц людей с использованием одного из статистических методов распознавания – скрытых марковских моделей (СММ) с дискретным временем [2]. Данный метод позволяет определять структурные особенности лица и учитывать характер искажений лица. Отметим, что скрытые марковские модели успешно применяются для обработки аудио сигналов, речи, изображений, распознавания текста, в биомедицине и других областях [3].

Мы остановимся на рассмотрении СММ с одномерной топологией. Несмотря на то, что этот тип моделей является простейшим, предыдущими исследователями (например, [2], [4]), как нам представляется, не до конца был раскрыт потенциал таких СММ.

2. СКРЫТЫЕ МАРКОВСКИЕ МОДЕЛИ

Далее отметим основные положения теории СММ и рассмотрим практические аспекты задачи распознавания лиц при помощи данного метода.

2.1 Элементы СММ

Элементами СММ являются:

1. Конечное множество скрытых состояний $S = \{s_1, s_2, \dots, s_N\}$.
2. Конечное множество наблюдаемых состояний (или дискретный алфавит) $V = \{v_1, v_2, \dots, v_M\}$.
3. Матрица переходных вероятностей $A = \{a_{ij}\}$, где $a_{ij} = P(q_t = S_j | q_{t-1} = S_i)$, $i, j = \overline{1, N}$, $t = \overline{1, T}$ ассоциируема со стационарной марковской цепью на пространстве скрытых состояний. Здесь q_t – состояние модели в момент времени t .
4. Начальная вероятность состояний $\pi = \{\pi_i\}$, $\pi_i = P(q_1 = S_i)$, $i = \overline{1, N}$.
5. Матрица эмиссий $B = \{b_i(k)\}$, $b_i(k) = P(o_t = v_k | q_t = s_i)$, $i = \overline{1, N}$, $k = \overline{1, M}$, o_t – наблюдаемый символ в момент времени $t = \overline{1, T}$.

Используя полунепрерывные СММ (непрерывное пространство наблюдаемых состояний, но дискретное пространство

скрытых состояний) имеем: $b_i(t) = \sum_{m=1}^{M_i} c_{im} N(o_t, \mu_{im}, \Sigma_{im})$,

где c_{im} – коэффициент k -ой смеси i -ого состояния, M_i – число наблюдаемых состояний, описывающих i -ое ненаблюдаемое состояние.

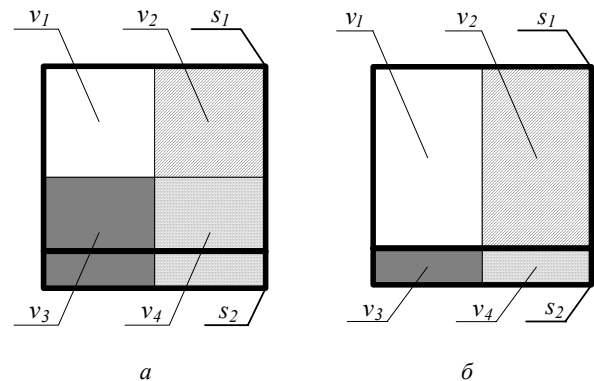
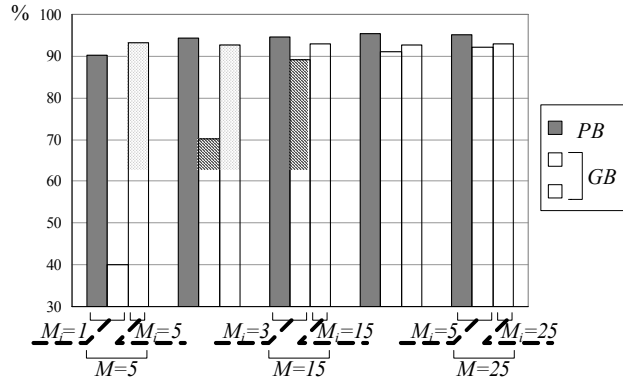


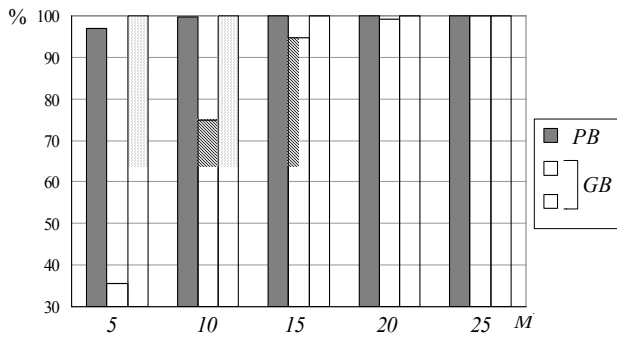
Рис. 1. Выбор наблюдаемых состояний в методе GB (a) и в методе PB (б)

Начиная с [5], где впервые было предложено использовать СММ для задачи распознавания, и в дальнейших работах на эту тему [2], [4], наблюдаемые состояния выбирались во всем пространстве наблюдений и использовались M или M_i

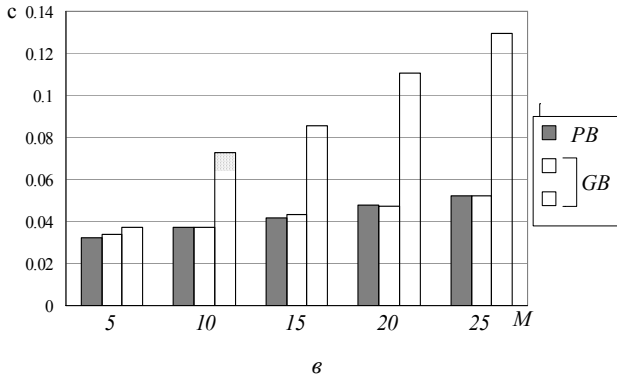
элементов для описания каждого скрытого состояния. Обозначим данный метод как *GB* (Global Bases).



а



б



в

Рис. 2. Зависимость процента распознавания по тестовой выборке (а), обучающей (б) и времени распознавания одного изображения (в) от M_i и M для модели *ID*, *UD*

Однако для наилучшего описания $b_i(t)$ мы предлагаем множество V выбирать таким образом, чтобы только M_i конкретных наблюдаемых состояний описывали скрытое состояние s_i (*PB* – Private Bases). Кроме того, эти наблюдаемые состояния выбираются из наблюдений, находящихся в этом скрытом состоянии.

Для описанных методов на рис. 1 схематично изображены наблюдаемые v_i , $i = \overline{1,4}$ и скрытые s_1, s_2 состояния.

Если положить $M_i = 2$, $i = \overline{1,2}$, то для описания $b_1(t)$, $t = \overline{1, T}$ будут использоваться наблюдаемые состояния v_1 и v_2 ; при этом в методе *PB* описание s_1 будет происходить намного точнее, чем в *GB*. Если же мы будем в методе *GB* использовать $M_i = M$, то, естественно, описание s_1 будет полное.

При одинаковом значении числа смесей M_i метод *PB* позволяет повысить процент распознавания (рис. 2, а) в сравнении с методом *GB*. Сравнивая метод *PB* и *GB*, использующий $M_i = M$, можно сделать вывод, что по качеству распознавания они практически идентичны, однако у метода *PB* скорость распознавания одного изображения выше (рис. 2, в). При малом значении $M_i \neq M$ метод *GB* недостаточно хорошо описывает обучающие данные (рис. 2, б). Предложенный же метод *PB* более точно настраивает СММ на воспроизведение обучающих наблюдений. В [6] мы показали, что вероятность генерации ложных наблюдений в методе *PB* с увеличением M_i снижается. При этом вероятность генерации своих тестовых наблюдений снижается. Таким образом, данный метод выбора наблюдаемых состояний приводит к высокому проценту и относительно высокой скорости распознавания.

2.2 Марковские цепи

Марковские цепи, используемые в работе, имеют одномерную топологию. На рис. 3 приведены примеры марковских цепей, а также обозначения соответствующих моделей.

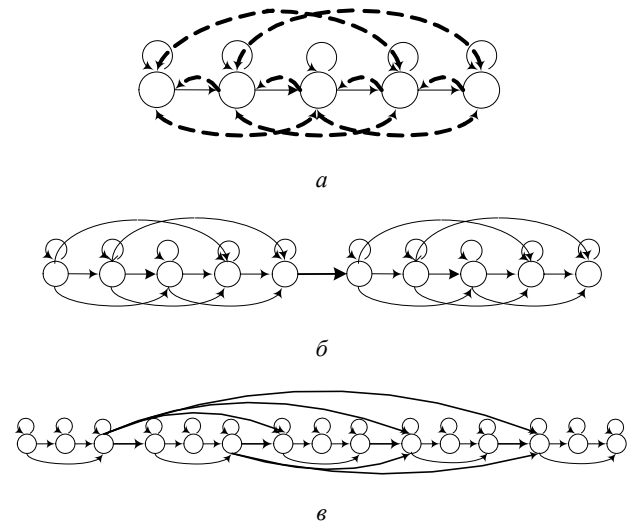


Рис. 3. Модель *ID* (а); модель *ID_TD* (б); модель *ID_2D* (в)
Условно каждое состояние ассоциируется с какой-либо частью лица. Например, состояния цепи для *ID* модели можно интерпретировать как лоб, глаза, нос, губы, подбородок. Этот тип модели является простейшим, используемым для распознавания лиц [2]. Модель *ID_2D* близка к псевдо-двумерным СММ, а *ID_TD* (*ID Two Domain*) предложена нами как некий переходный вариант между *ID* и *ID_2D*.

Нами было установлено, что модель *ID_TD* сочетает в себе достоинство *ID_2D* модели – высокий процент распознава-

ния и приемлемую скорость распознавания, присущую $1D$ модели [7].

На рис. 3, а наличие связей, обозначенных пунктирной линией, регулируется некоторыми дополнительными параметрами модели (см. пункт 3.2). Наличие дополнительных переходов возможно также и для моделей $1D_TD$ и $1D_2D$.

2.3 Этапы обучения и распознавание СММ

Входные данные для системы распознавания – это наборы изображений лиц, соответствующие различным людям. Каждой персоне ставится в соответствие СММ, подбор параметров которой происходит в процессе обучения. На этом этапе каждая модель $\lambda_r = (\pi_r, A_r, B_r)$, $r = \overline{1, R}$ (R – число персон) настраивается на свои обучающие изображения в результате максимизации функций $P(O^{(r)}, Q | \lambda_r)$ и $P(O^{(r)} | \lambda_r)$. Решение первой подзадачи, т.е. определение максимизирующей последовательности скрытых состояний при фиксированной последовательности наблюдаемых состояний, дает алгоритм динамического программирования Витерби. Модификация EM алгоритма (алгоритм Баум–Уэлша) используется для решения второй подзадачи. В результате обучения полученная модель λ_r может быть использована в качестве генератора наблюдаемой последовательности $O^{(r)}$ (извлеченных из обучающих изображений), а также при распознавании сигналов (тестовых изображений) в некотором смысле близких к $O^{(r)}$. Тестовое изображение считается относящимся к персоне p , если выполняется:

$$P(O^{(test)} | \lambda_p) = \max_{r=1, R} P(O^{(test)} | \lambda_r).$$

Все исследования, представленные в данной статье, проводились на базе изображений ORL^1 . Обучение каждой модели происходило по 5 случайно выбранным фотографиям персоны и их преобразованиям (см. пункт 3.3.2). Распознаванию подвергались не вошедшие в обучающий набор изображения. Результаты усреднялись по 10 выборкам.

3. НАБЛЮДЕНИЯ

Цифровое изображение возможно интерпретировать как случайный двумерный дискретный сигнал, который наблюдается системой. Последовательность наблюдений $O = o_1 o_2 \dots o_T$, где o_t – сигналы, упорядоченные по про-

¹ ORL – Olivetti Research Ltd. Данная база состоит из 400 изображений 40 людей – сотрудников Olivetti и студентов Кембриджского университета, по 10 полутоновых изображений на каждого человека; размер изображений 92×112 пикселей; база содержит изображения лиц с различной эмоциональной окраской, с очками/ без очков, разные повороты голов; возраст людей: от 18 до 81 года, в основном между 20 и 35 годами; в базе данных содержится 4 женских лица и 36 мужских; съемка проводилась в различное время, с разной освещенностью, но всегда с темным фоном.

В этой статье используется изображение одной из персон, принадлежащей базе ORL.
<http://www.orl.co.uk/face database.html>

странственным отношениям, может извлекаться из изображения различными способами. В силу этого описательные способности полученных моделей могут различаться.

3.1 Извлечение наблюдений

Мы использовали окна сканирования изображения размерами $X \times L_y$ (сканирование по изображению – сверху вниз), $L_x \times Y$ (слева на право), $L_x \times L_y$ (слева на право, сверху вниз) (см. рис. 4). Обозначим эти методы сканирования соответственно UD , LR , $LRUD$. В результате получается последовательность частей изображения, к которой с целью уменьшения размерности пространства наблюдений и снижения вычислительной сложности применяются некоторые преобразования (например, дискретное косинусное преобразование). На рис. 5 приведен фрагмент такой полученной последовательности (сканирования методом $LRUD$).

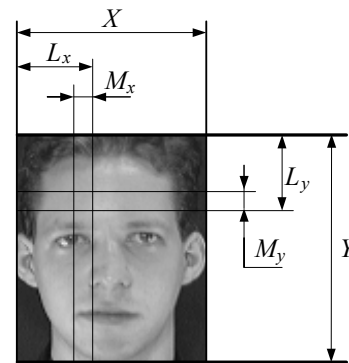


Рис. 4. Геометрические параметры изображения

Число наблюдений, извлеченных из одного изображения методом UD , определяется как $T_Y = \frac{Y - L_y}{L_y - M_y} + 1$. Общее

число наблюдений, извлеченных из обучающего набора Z изображений для одной модели: $T = T_X \cdot T_Y \cdot Z$. Далее в пункте 3.3 мы рассмотрим проблему выбора числа наблюдений.

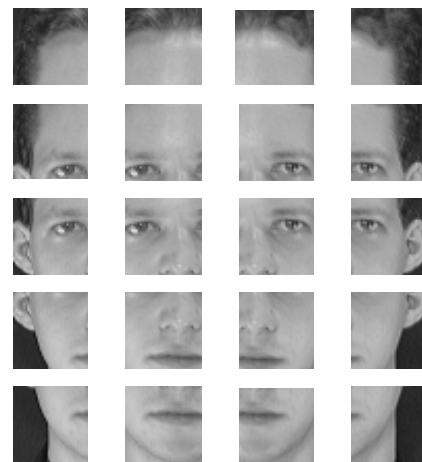


Рис. 5. Последовательность частей изображения

Извлечение наблюдений для ID модели происходит методом UD либо $LRUD$; ID_TD – для первой подмодели UD , для второй – LR или UD (параметры окна сканирования выбираются иные, чем для первой подмодели); ID_2D – $LRUD$.

3.2 Ограничения

С целью придания моделям дополнительной гибкости и уменьшения вычислительной сложности на данные были наложены некоторые ограничения по пребыванию наблюдений в скрытых состояниях. На рис. 6 на примере ID модели иллюстрируется данная идея.

Наблюдения, извлеченные из области A , могут находиться в S_i скрытом состоянии, в $B - S_{i+1}$, в области C – как в S_i , так и S_{i+1} . Регулируя глубины проникновения наблюдений в соседние состояния, мы изменяем число различных переходов в марковской цепи. Например, в случае, если $G_{down} > B$, то переход возможен из состояния S_i в S_{i+2} . Таким образом, имеем набор дополнительных параметров: $G = \{G_{up}, G_{down}, G_{left}, G_{right}\}$.

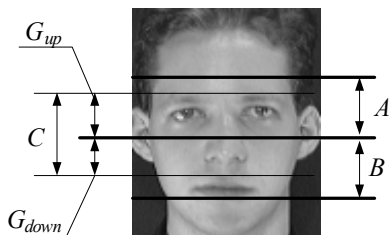


Рис. 6. Дополнительные параметры модели

В предельном случае мы получаем эргодическую марковскую цепь, однако такая модель из-за высокой степени свободы плохо настраивается на обучающие данные, что негативно сказывается на качестве распознавания.

3.3 Выбор числа наблюдений

Выбор числа наблюдений T влияет на:

1. Способность модели адекватно описывать обучающие данные.
2. Способность модели различать тестовые данные.
3. Скорость выполнения этапов обучения и распознавания.
4. Объем памяти, используемый программным приложением.

При малых объемах используемых баз данных изображений и современном уровне техники четвертую проблему будем считать несущественной. Таким образом, следует найти компромиссный вариант между первыми двумя пунктами и скоростью распознавания (скорость обучения также не будем учитывать). При слишком большом значении T возникает проблема переобученности моделей; при недостаточно большом – невозможность адекватно различать модели.

3.3.1 Условия выбора числа наблюдений

Рассмотрим некоторые дополнительные условия, которые помогут обозначить диапазон изменения значения T .

Условие 1.

Из теории планирования эксперимента известно, что для оценки q параметров модели необходимо не более $\frac{q(q+1)}{2} + 1$ точек плана эксперимента.

Для каждого скрытого состояния S_i , $i = \overline{1, N}$ и наблюдаемого состояния v_m , $m = \overline{1, M_i}$ надо оценить вектор математического ожидания и ковариационную матрицу: μ_{im}, Σ_{im} . Заметим, что мы используем в работе диагональную ковариационную матрицу – это дает большую обобщающую способность моделям и уменьшает вычислительную сложность этапов обучения и распознавания. Таким образом, необходимо оценить $2M$ векторов. Учитывая все выше сказанное, получаем, что $T_{\min} = M \cdot (2M + 1) + 1$ наблюдений должно быть извлечено для обучения одной модели (метод PB).

Мы не случайно выбрали обозначение T_{\min} , так как необходимо, чтобы около $2M + 1$ наблюдений находилось в каждом наблюдаемом состоянии.

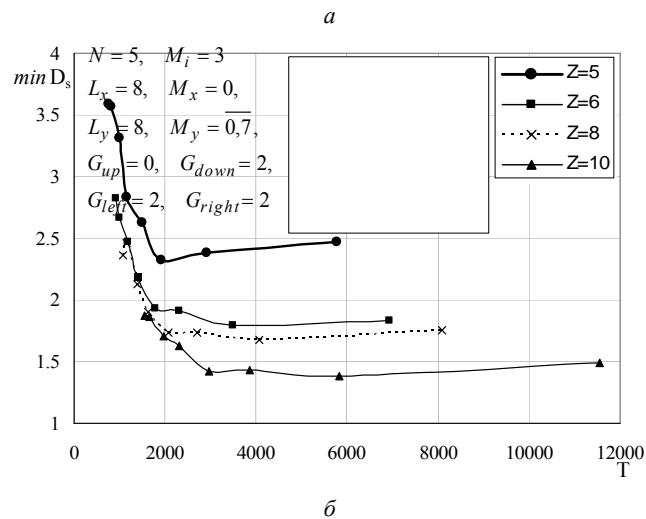
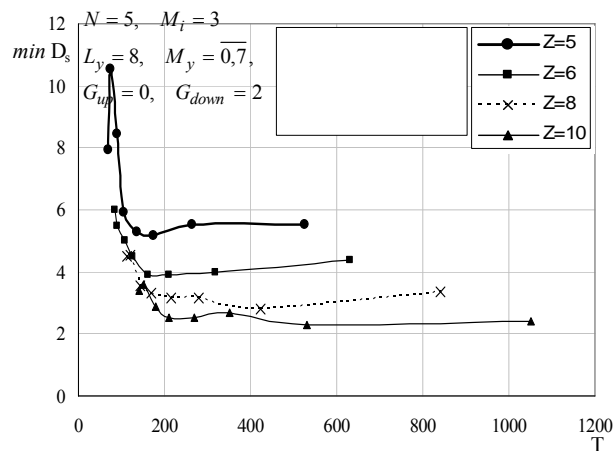


Рис. 7. Зависимость минимального расстояния от числа наблюдений в обучающей выборке для модели ID , UD (а) и ID_2D , $LRUD$ (б)

При использовании *PB* моделей с $N = 5$, $M_i = 3$, $i = \overline{1, N}$, получаем, что $T_{\min} = 466$.

Условие 2.

Необходимо иметь возможность сравнения моделей. Измерение расстояния между моделями было предложено в [8]; данная мера основана на расстоянии Каллбэка-Либлера:

$$D_s(\lambda_1, \lambda_2) = \frac{1}{2} [D(\lambda_1, \lambda_2) + D(\lambda_2, \lambda_1)], \text{ где}$$

$$D(\lambda_1, \lambda_2) = \lim_{T \rightarrow \infty} \frac{1}{T} [\ln P(O^1 | \lambda_1) - \ln P(O^1 | \lambda_2)].$$

Таким образом, необходимо найти некоторое значение T , начиная с которого $D_s(\lambda_1, \lambda_2)$ перестанет значительно изменяться. Расстояния будем смотреть по обучающей выборке.

На рис. 7 приведена зависимость $\min_{i,j} D_s(\lambda_i, \lambda_j)$, $i, j = \overline{1, R}$ от числа наблюдений.

Начиная с некоторого значения T количество информации, содержащееся в одном наблюдении, перестает существенно изменяться, то есть наблюдения не добавляют новых сведений об обучающих данных. Таким образом, эффективность увеличения числа наблюдений T падает.

Условие 3.

Этап распознавания одного изображения имеет вычислительную сложность порядка $O(T \cdot N)$.

3.3.2 Увеличение объема обучающей выборки

Существует два пути увеличения числа наблюдений:

1. Увеличение числа обучающих изображений.
2. Изменение параметров сканирующего окна (это более естественное решение увеличения числа наблюдений для реальной задачи распознавания).

Объем обучающей выборки ограничен: в базе данных присутствует Z_r фотографий r -ой персоны, где $r = \overline{1, R}$. Различными преобразованиями можно получить дополнительный набор обучающих изображений. Следует иметь в виду, что изображения из этого набора должны в некоторой степени «предсказывать» тестовые изображения.

Мы использовали простейшее преобразование над обучающими изображениями – зеркальное отражение.

Приведенные графики на рис. 7, 8 соответствуют различному числу изображений, присутствующих в обучающем наборе для каждой персоны. В каждом эксперименте с различным числом обучающих изображений изменялись параметры сканирующего окна.

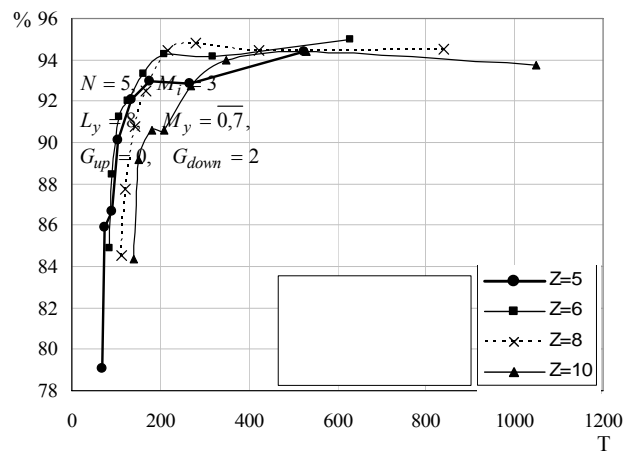


Рис. 8, а. Зависимость процента распознавания от числа наблюдений в обучающей выборке для модели *ID*, *UD*

При $Z = 6$ в наборе было одно зеркально-отраженное изображение случайно выбранной фотографии. Введение дополнительных наблюдений в обучающий набор позволяет повысить точность распознавания, с сохранением постоянного времени распознавания.

Заметим, что различные модели при одном и том же числе наблюдений имеют различный процент распознавания (рис. 8, а, б). Этот аспект был исследован нами в [7].

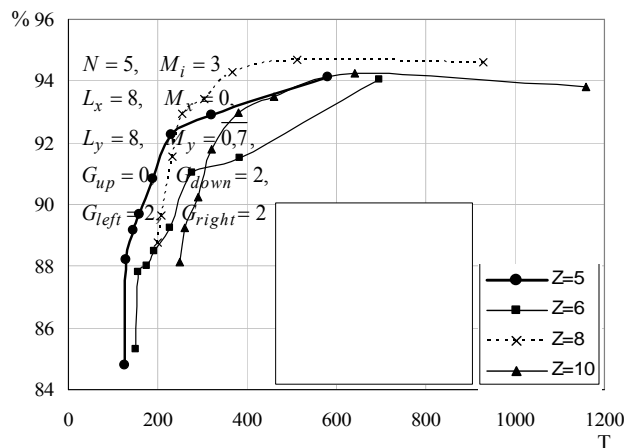


Рис. 8, б. Зависимость процента распознавания от числа наблюдений в обучающей выборке для модели *ID*, *TD*, *UD*, *LR*

Обратим внимание, что характеристики окна сканирования влияют не только на число наблюдений, но и на распознавательную способность модели. Рис. 8, б аналогичен рис. 8, а, но модель используется с другими параметрами окна. При одном и том же числе наблюдений (и при одинаковых значениях Z) проценты распознавания различаются.

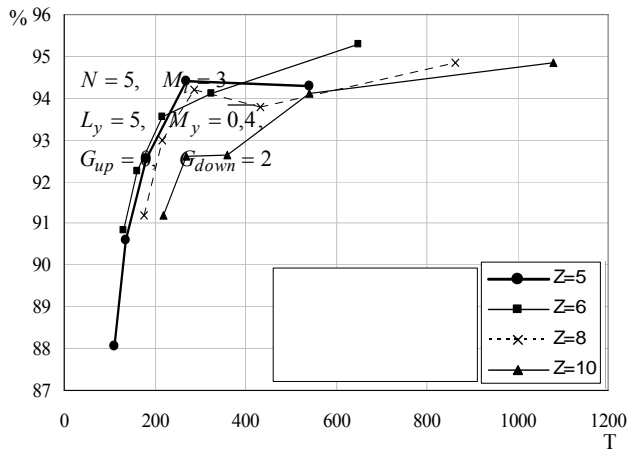


Рис. 8, в. Зависимость процента распознавания от числа наблюдений в обучающей выборке для модели $1D$, UD

Отметим, что выход на максимально возможное для модели значение процента распознавания достигается так же при отношении параметров сканирующего окна: $\frac{M}{L} \approx 0.75$.

Данный результат аналогичен приведенному в [2].

4. ЗАКЛЮЧЕНИЕ

В данной статье были предложены некоторые новые методы выбора элементов СММ, влияющих на описательную способность модели (такие как, наблюдаемые состояния, марковские цепи). Рассмотрен вопрос выбора способа извлечения наблюдений и их числа.

Проведенные сравнения (например, в [9]) показали, что методы, основанные на более сложных СММ, достигают ошибки распознавания, близкой к нулю, что является наилучшим результатом среди рассмотренных методов распознавания лиц (на базе *ORL*). Рассмотренные нами СММ с одномерной топологией уступают по точности псевдо-двумерным и СММ, использующим иные пространства признаков. Однако, предложенные изменения в описании скрытых наблюдений и переход от $1D$ модели к более сложным моделям с одномерной топологией, позволили увеличить процент распознавания в среднем на 10% (сравнить с [9]).

В дальнейшем мы предполагаем заняться разработкой способа извлечения наиболее информативных наблюдений из изображений с привлечением методов планирования эксперимента.

5. БИБЛИОГРАФИЯ

- [1] Д. Брилюк, В. В. Старовойтов. *Распознавание человека по изображению лица и нейросетевые методы*// Обзорный препринт: <http://daily.sec.ru/dailytblshow.cfm?id=5&pid=4425&pos=1&stp=50>
- [2] F. S. Samaria. *Face Recognition Using Hidden Markov Models*// PhD Thesis, University of Cambridge, 1994.
- [3] O. Cappe. *Ten years of HMM*// CNRS, LTCI & ENST, Dpt. TSI <http://www.tsi.enst.fr/~cappe/>
- [4] A. V. Nefian, M. H. Hayes III. *Hidden Markov Models For Face Recognition*// ICA SSP98, 1998, vol. 5, pp 2721-2724.

- [5] L. R. Rabiner. *A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition*// Proceedings of the IEEE, 1989, vol. 77, no. 2, pp 257-285.
- [6] Т. А. Гульятеева, А. А. Попов. *Скрытые марковские модели с одномерной топологией в задаче распознавания лиц*// Сборник научных трудов НГТУ. – 2006. – №1 (в печати).
- [7] Т. А. Гульятеева, А. А. Попов. *Применение скрытых марковских моделей к задаче распознавания лиц*// Труды VI международной научной конференции «Наука и образование» – 2006 (в печати).
- [8] В. Н. Juang, L. R. Rabiner. *A Probabilistic Distance measure between HMMs*// AT&T Tech., Feb. 1985, vol. 64, no. 2, pp 391-408.
- [9] M. Bicego, U. Castellani, V. Murino. *Using hidden Markov models and wavelets for face recognition*// sci.univr.it/~bicego/iciap2003.pdf

Об авторах

Татьяна Александровна Гульятеева – магистр факультета прикладной математики и информатики Новосибирского государственного технического университета, преподаватель-стажер кафедры программных систем и баз данных Новосибирского государственного технического университета

Адрес: Новосибирск, 630092, пр-т Карла Маркса, 20, НГТУ, 1-ый учебный корпус, факультет ПМИИ, кафедра ПСиБД.

Телефон: (383) 3460600

E-mail: gtany@mail.ru

Александр Александрович Попов – профессор, доктор технических наук, заведующий кафедры программных систем и баз данных Новосибирского государственного технического университета

Адрес: Новосибирск, 630092, пр-т Карла Маркса, 20, НГТУ, 1-ый учебный корпус, факультет ПМИИ, кафедра ПСиБД.

Телефон: (383) 3460600

E-mail: alex@fpm.ami.nstu.ru

Modifications 1D hidden markov models for problem of person recognition

Abstract

In this paper one of statistical methods of face recognition, based on the one-dimensional Hidden Markov Models represented. Some modifications of models increasing recognizing properties of a system are considered.

Keywords: *Hidden Markov Models face recognition.*

About the authors

Tatyana A. Gulyaeva is a M.S. student at Novosibirsk State Technical University, Department of Software and Database Engineering. Hers contact email is gtany@mail.ru

Alexander A. Popov is a professor at Novosibirsk State Technical University, head of Department of Software and Database Engineering. His contact email is alex@fpm.ami.nstu.ru