

ОБРАБОТКА ИЗОБРАЖЕНИЙ, РАСПОЗНАВАНИЕ ОБРАЗОВ

МЕТОД БЫСТРОЙ КОРРЕЛЯЦИИ С ИСПОЛЬЗОВАНИЕМ ТЕРНАРНЫХ ШАБЛОНОВ ПРИ РАСПОЗНАВАНИИ ОБЪЕКТОВ НА ИЗОБРАЖЕНИЯХ

Н.И. Глузов¹, Е.В. Мясников^{1,2}, В.Н. Копенков^{1,2}, М.А. Чичева¹

¹ Институт систем обработки изображений РАН, Самара, Россия,

² Самарский государственный аэрокосмический университет имени академика С.П. Королёва, Самара, Россия

Аннотация

В работе рассматривается задача поиска и распознавания на изображениях фрагментов, соответствующих одному из множества шаблонов. Предлагается метод быстрой корреляции по множеству тернарных шаблонов, который успешно разрешает перечисленные проблемы. Продемонстрировано его применение в двух задачах анализа изображений: поиска положения глаз на документальных фотографиях лиц и распознавания машиночитаемых строк на сканированных изображениях документов.

Ключевые слова: корреляция, распознавание объектов, изображения документов, поиск глаз, машиночитаемые строки.

Введение

В работе рассматривается задача поиска и распознавания на изображениях фрагментов, соответствующих одному из множества шаблонов. Такая задача возникает, например, при распознавании текста [5,7]. Другим примером является поиск характерных точек на изображениях лиц, таких как координаты зрачков, уголки губ и т.п. [4,6]. Эти задачи похожи тем, что искомым образцом может принимать существенно различный вид, при этом следует определить не только его местоположение, но и меру сходства с каждым из образцов с целью принятия окончательного решения.

Согласно общепринятому подходу к решению таких задач вычисляется корреляция входного изображения с каждым из шаблонов, после чего анализируются полученные корреляционные поля [2]. Независимо от того, какой способ вычисления корреляции будет выбран (прямое вычисление свертки, через дискретное преобразование Фурье и т.п. [3]), такой подход требует решения ряда проблем. Первая из них – это высокая вычислительная сложность, которая естественно возрастает с ростом числа шаблонов. Вторая – поиск критерия, который позволит из множества точек корреляционных полей выбрать наиболее соответствующую истинному значению. Наконец, в связи с тем, что на реальных изображениях объекты, как правило, находятся на сложном фоне, искажены, зашумлены, необходимо уменьшить влияние мешающих факторов.

К настоящему времени существует ряд методов, в той или иной мере решающих изложенные проблемы. Так, использование параллельно-рекурсивной реализации при помощи аппроксимации шаблонов специальным базисом [2] позволяет существенно снизить время обработки. Упрощение шаблонов (например, бинаризация), изменение формы учитываемой области шаблона (уход от прямоугольного «окна») позволяет как частично решить проблему быстродействия, так и уменьшить влияние мешающих факторов. Специальная обработка кор-

реляционных полей может облегчить поиск и расчет критериев. Однако ни один из названных методов в чистом виде не позволяет эффективно решить поставленную задачу. Ниже предлагается метод, сочетающий в себе перечисленные подходы.

1. Метод быстрой корреляции

Предлагаемый метод быстрой корреляции с предварительно подготовленным множеством шаблонов состоит из следующих шагов.

1. *Предварительные действия* включают в себя подготовку набора шаблонов (выполняется один раз для всего набора изображений).

2. *Определение области*, в которой будет производиться сравнение с шаблонами, может быть выполнено как оператором вручную, так и автоматически при помощи специально разработанных алгоритмов.

3. *Формирование корреляционных полей* заключается в расчете корреляции с каждым из шаблонов для всех его положений в области поиска.

4. *Обработка корреляционных полей* выполняется с целью отбора перспективных точек. При этом учитываются как очевидные соображения (например, что два обнаруживаемых объекта могут располагаться на расстоянии не менее своего размера), так и специальные, определяемые спецификой задачи. В результате этого этапа на каждом из полей остается небольшое количество точек, которые анализируются на последнем этапе.

5. *Принятие решения* осуществляется на основе анализа корреляционных полей в области выбранных точек. Необходимо по набору точек корреляционных полей выбрать координаты искомого объекта. Простейшим критерием выбора является максимальное значение корреляции по всем оставшимся точкам всех корреляционных полей.

Подготовка шаблонов

В рассматриваемой задаче к шаблонам предъявляются противоречивые требования. С одной сторо-

фики используемых в машиночитаемых строках шрифтов (в частности, различная толщина линий символов на документах различных стран). Пример шаблонов цифр показан на рис.3.



Рис.3. Шаблоны символов «цифра»

Шаблоны подготавливаются в виде матриц $T^{(m)}(v, h)$, $0 \leq m < M$, $0 \leq v < V_s$, $0 \leq h < H_s$, где M - количество шаблонов, V_s, H_s - размеры шаблона. При определенной знакопозиции (v_0, h_0) распознаваемого символа корреляционные поля $\delta^{(m)}(i, j)$ вычисляются в области $v_0 - \Delta \leq i \leq v_0 + \Delta$, $h_0 - \Delta \leq j \leq h_0 + \Delta$, где Δ - априорно заданное значение. Для каждого поля выбирается максимальное значение корреляции. По максимальным значениям для различных корреляционных полей формируется список шаблонов, наиболее похожих на текущий распознаваемый символ.

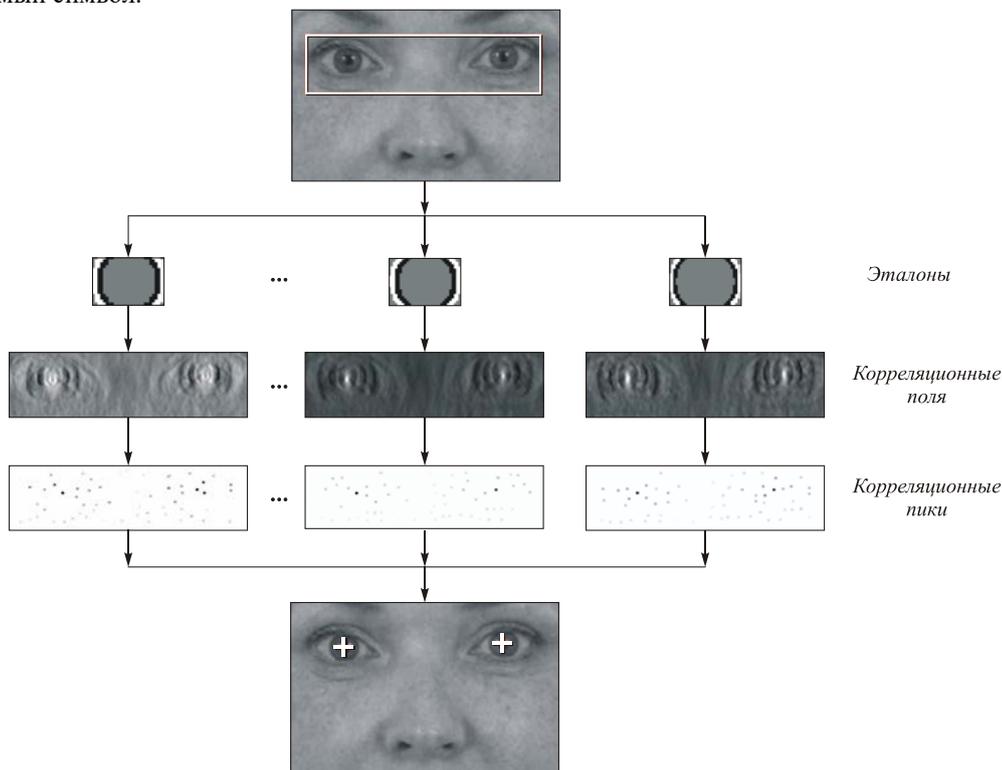


Рис.4. Схема алгоритма поиска центров роговиц глаз

Для поиска глаз формируются тернарные шаблоны, отличающиеся радиусом роговицы (рис.5). Легко видеть, что для таких шаблонов в расчете свертки участвует небольшое количество отсчетов. Это обеспечивает высокую скорость вычисления.



Рис.5. Пример шаблона для поиска центров роговиц

Таким образом, после распознавания всех символов машиночитаемой строки формируется матрица решений $R(n, k)$, на основе которой окончательно формируется распознанная машиночитаемая строка.

Произведен статистический эксперимент на 587 изображениях загранпаспортов 34 стран. Вероятность верного распознавания отдельного символа составила 0,998, вероятность верного распознавания всей машиночитаемой строки – 0,982.

Задача поиска положения глаз на документальных фотографиях лиц

Эта задача возникает в рамках практически всех алгоритмов распознавания лиц на изображениях, первым шагом в которых выполняется геометрическая нормализация с привязкой центров роговиц глаз к заранее заданным точкам. Предлагаемый алгоритм полностью соответствует общей схеме, описанной в разделе 1. Схематично он показан на рис.4.

С другой стороны, объем вычислений в значительной степени определяется размерами области расчета свертки. Для сокращения этой области разработан алгоритм построения «каркаса» лица, основанный на анализе поля локальных дисперсий изображения. Предварительно на поле дисперсий некоторого эталонного изображения были выбраны два набора опорных точек:

$$S^{(+)} : \left\{ \left(v_i^{(+)}, h_i^{(+)} \right) \right\} \text{ и } S^{(-)} : \left\{ \left(v_j^{(-)}, h_j^{(-)} \right) \right\},$$

соответствующих точкам с высокой локальной дисперсией (границы глаз, губ, нос и т.п.) и низкой локальной дисперсией (гладкие участки изображения с плавно меняющейся яркостью: щеки, лоб) соответственно. В качестве эталонного изображения в работе было использовано изображение, представляющее усредненное поле локальных дисперсий по базе из 4800 геометрически нормализованных фотографий (центры роговиц глаз находятся в фиксированных точках). Данное изображение и множества опорных точек показаны на рис. 6.

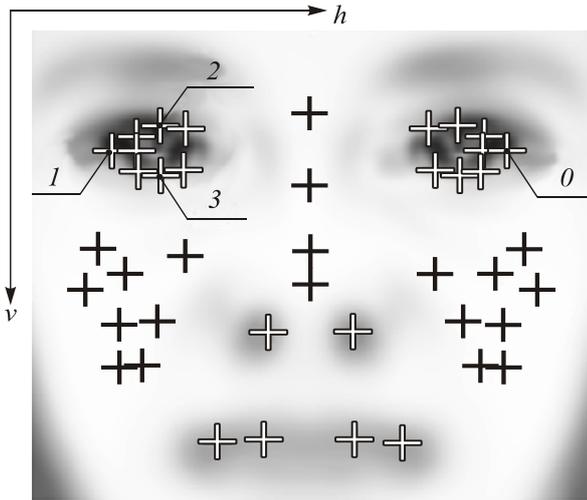


Рис.6. Фрагмент инвертированного усредненного поля локальных дисперсий с нанесенными на него опорными точками

Далее при обработке конкретного изображения эти наборы точек подвергаются независимому по осям линейному преобразованию координат:

$$V_i = av_i + b, \quad H_i = ch_i + d,$$

параметры которого определяются из условия оптимизации функционала:

$$F(a,b,c,d) = \sum_{S^{(+)}} g(V_i^{(+)}, H_i^{(+)}) - \sum_{S^{(-)}} g(V_j^{(-)}, H_j^{(-)}) \rightarrow, \\ \rightarrow \max_{a,b,c,d}$$

где $g(V_i^{(\pm)}, H_i^{(\pm)})$ - поле локальных дисперсий.

Далее по найденному положению «каркаса» (при оптимальных значениях параметров функционала) формируется зона поиска глаз.

Зона поиска глаз формируется как прямоугольник с размерами

$$\Delta H = h_0 - h_1 + 2 \cdot k_h \cdot H_{et}, \\ \Delta V = v_3 - v_2 + V_{et} \cdot (k_v^{up} + k_v^{down})$$

и началом в точках с координатами

$$H_0 = h_1 - k_h \cdot H_{et}, \\ V_0 = v_2 - k_v^{up} \cdot V_{et}.$$

Здесь точки $(v_0; h_0), (v_1; h_1), (v_2; h_2), (v_3; h_3)$ показаны на рис.6, V_{et}, H_{et} - размеры наибольшего эталона, $k_h, k_v^{up}, k_v^{down}$ - коэффициенты, подбираемые таким образом, чтобы центры роговиц попадали в зону поиска с требуемой вероятностью.

На рис. 7 показаны примеры изображений с найденными опорными точками (красные кресты) и отмеченными зонами поиска глаз (красные рамки). На рис. 7(б) положение каркаса смещено вверх, но глаза тем не менее входят в зону поиска.

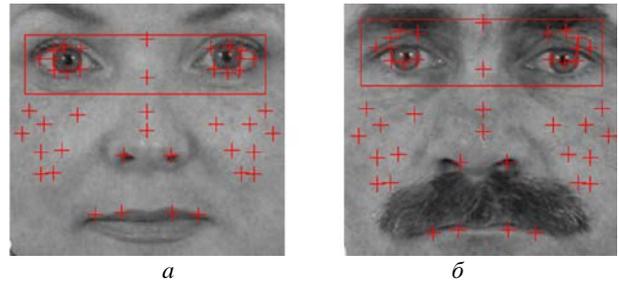


Рис.7. Фрагмент инвертированного изображения

Обработка (за исключением выбора опорных точек на эталонном изображении, выполняемого однократно и заранее) является полностью автоматической.

Далее определяются точные положения глаз, в качестве которых принимаются центры роговиц. Для их нахождения используется набор шаблонов в некотором диапазоне радиусов $[\bar{R}_{min}, \bar{R}_{max}]$, который определяется предварительно, исходя из априорной информации о разрешении.

Для каждого конкретного изображения диапазон рассматриваемых радиусов изображения выбирается в этих пределах следующим образом:

$$R_{min} = \frac{D_c}{\mu} - d_1 \geq \bar{R}_{min}, \\ R_{max} = \frac{D_c}{\mu} + d_2 \leq \bar{R}_{max}, \tag{1}$$

где $D_c = h_4 - h_2$ - полурасстояние между глазами, определенное по каркасу, в свою очередь h_2, h_4 - координаты точек каркаса по горизонтали (см. рис. 6); μ - отношение расстояния между глазами к некоторому усредненному радиусу роговицы, d_1, d_2 - разброс радиусов от среднего.

Если какое-либо из неравенств в (1) не выполняется, то принимается $R_{min} = \bar{R}_{min}$ или $R_{max} = \bar{R}_{max}$.

После расчета корреляционных полей по каждому из них строится поле локальных максимумов, или корреляционных пиков (см. рис. 4). Из них, в свою очередь, выбирается наиболее «перспективная» пара, которая, во-первых, удовлетворяет ряду ограничений, а во-вторых, обеспечивает максимум суммы корреляционных пиков. Упомянутые ограничения естественным образом вытекают из расположения глаз и требований к фотографиям на документы. Кроме того, ограничения используют информацию о положении центров роговиц, оценен-

ном с помощью опорных точек «каркаса». Рис.8 иллюстрирует смысл основных величин, используемых при проверке ограничений (точками обозначены «кандидаты» на центры роговиц, крестами - центры роговиц, найденные с помощью построения «каркаса»).

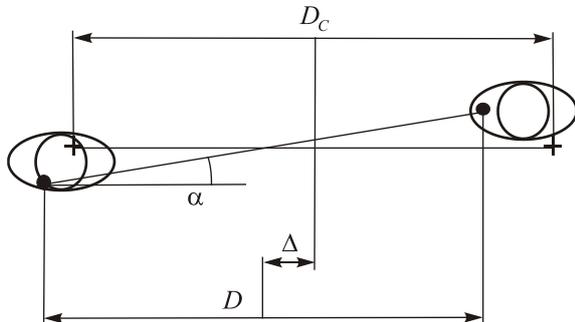


Рис.8. Иллюстрация ограничений, накладываемых на «кандидатов» на центры роговиц глаз

Расстояние между «кандидатами» на центры роговиц D по горизонтали должно лежать в определенных границах D_{\min} и D_{\max} , определяемых по текущему радиусу роговицы.

Модуль угла наклона линии глаз α , отсчитываемый от горизонтали, не должен превышать определенную величину: $|\alpha| \leq \bar{\alpha}$ (обычно $5^\circ-7^\circ$).

«Кандидаты» на центры роговиц должны быть симметричны относительно центра лица (определяемого как вертикальная ось симметрии области поиска глаз) с определенной точностью (ограничение на величину отклонения Δ задается текущим радиусом роговицы).

Отношение расстояния по горизонтали между «кандидатами» на центры роговиц D к расстоянию между центрами глаз, определенными с помощью алгоритма построения «каркаса» D_c , должно лежать в определенных границах.

Для каждой точки из пар, удовлетворяющих ограничениям, вычисляется значение критерия:

$$J = \frac{\max - \min}{E - \min}, \quad (2)$$

где \max , \min , E - соответственно максимальное, минимальное и среднее значения корреляционных полей внутри определенной области с центром в текущей точке.

Окончательные точки, соответствующие положению центров роговиц глаз, выбираются из условий:

- пара точек удовлетворяет вышеупомянутым ограничениям;
- разница радиусов роговиц одной пары глаз не превышает одного пикселя;
- суммарное значение критерия (2) для выбранной пары максимально.

Экспериментальное исследование алгоритма поиска положения глаз

Экспериментальное исследование алгоритма поиска глаз проводилось на базе, содержащей около 5000 изображений. Это фотографии лиц, предназначенные для использования на документах. С одной стороны, эти фотографии отвечают всем требованиям к документальным фото, с другой - это качественные цифровые фотографии, выполненные в одинаковых условиях. Высокое качество, безусловно, облегчает задачу поиска, зато второе свойство - однородность условий - делает эту базу удобной для исследований.

На рис. 9 показана вероятность P попадания центров роговиц глаз в область поиска S в зависимости от относительной площади области, где

$$P = P\left(\left(\bar{v}_1, \bar{h}_1\right), \left(\bar{v}_2, \bar{h}_2\right) \in S\right),$$

$\left(\bar{v}_1, \bar{h}_1\right), \left(\bar{v}_2, \bar{h}_2\right)$ - истинные положения центров роговиц, S - область с началом в точке (V_0, H_0) , размером $(\Delta V, \Delta H)$, описанная выше и показанная на рис. 10, $s = \frac{\Delta V \Delta H}{VH}$ - относительная площадь области поиска.

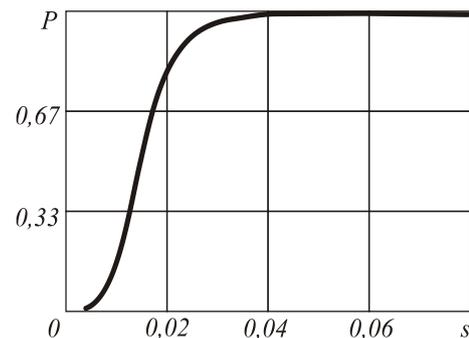


Рис.9. Вероятность попадания центров роговиц глаз в область поиска в зависимости от относительной площади области

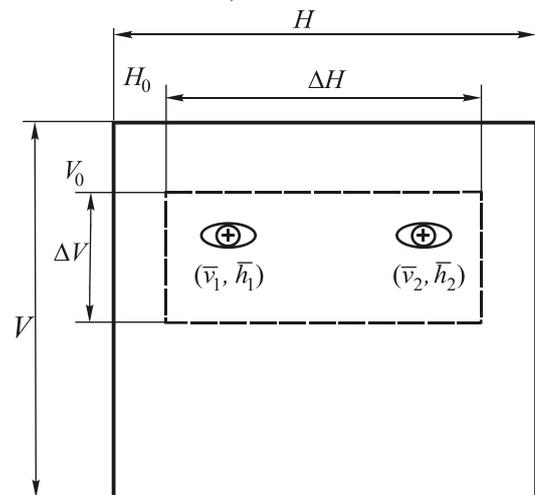


Рис. 10. Область поиска центров роговиц

Из рис. 10 видно, что глаза попадают в зону поиска, площадь которой не превышает 4% от площади

ди изображения, с вероятностью близкой к единице ($P=0,998$).

На рис.11 показана интегральная функция распределения вероятности отклонения найденных центров роговиц глаз от их истинных положений:

$$P(\max(\rho_1, \rho_2) \leq l),$$

где $\rho_i = \sqrt{(\bar{v}_i - v'_i)^2 + (\bar{h}_i - h'_i)^2}$ - расстояния между истинным и найденным положением центров роговиц, $(v'_1, h'_1), (v'_2, h'_2)$ - найденные алгоритмом положения центров роговиц (см. иллюстрацию на рис. 12).

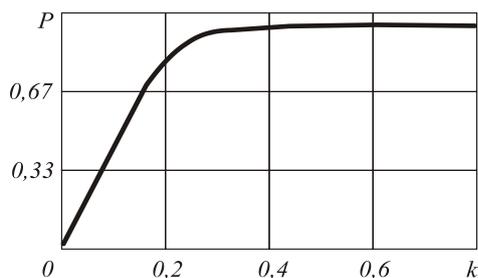


Рис.11. Интегральная функция распределения вероятности относительного отклонения найденных центров роговиц глаз от их истинных положений

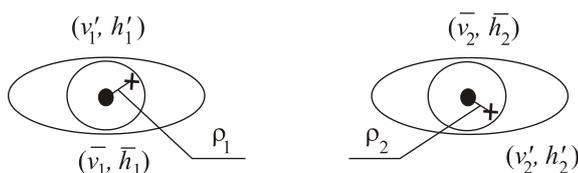


Рис. 12. Определение расстояния до истинных центров роговиц

Величина l задается в пикселах, однако фотографии, включенные в эту базу, имеют разные размер и разрешение. Поэтому для однородности критерия будем оценивать величину относительного отклонения найденных центров роговиц глаз от истинных положений:

$$k = \frac{l}{r},$$

где r - истинный радиус роговицы. То есть значение коэффициента $k=0$ означает точное попадание в центр роговицы, значение $k=1$ - попадание на край роговицы.

Из графика видно, что с вероятностью 0,95 найденные точки отличаются от истинных не более чем на половину радиуса роговицы, то есть заведомо лежат в пределах зрачка глаза.

Заключение

Таким образом, в работе предложен метод быстрой корреляции с использованием множества шаблонов и показано его применение к двум задачам анализа изображений: поиска положения глаз на документальных фотографиях лиц и распознавания машиночитаемых строк на сканированных изображениях документов. По сравнению с традиционным подходом значительно возрастает скорость обработки, что позволяет провести сравнение с большим числом шаблонов, тем самым повысив качество анализа.

Благодарности

Работа выполнена при поддержке российского фонда фундаментальных исследований (РФФИ), проекты № 06-01-00722, 06-01-00616, 07-01-96612, 07-07-97610, 08-07-90704-моб_ст, в рамках российско-американской программы «Фундаментальные исследования и высшее образование» (CRDF Project RUX0-014-SA-06) и гранта Президента РФ по поддержке ведущих научных школ (НШ-3086.2008.9).

Литература

1. Документ ICAO 9303 "Machine Readable Travel Documents(MRTD)" //http://www.icao.int/mrtd/publications/doc.cfm
2. **Сойфер, В.А.** Методы компьютерной обработки изображений / В.А. Сойфер, - М.:Физматлит, 2001. - 784с.
3. **Chicheva M.A.** Optimization of linear filtering procedure in image processing applications. / M.A. Chicheva, N.I. Glumov, V.V. Sergeev //Proceedings of the 6-th German-Russian workshop "Pattern recognition and image understanding" OGRW-6-2003, 2003. - P.149-154.
4. **Kawato, S.** Two-step approach for real-time eye tracking with a new filtering technique / S. Kawato; J. Ohya //Systems, Man, and Cybernetics, 2000. - Vol.2. - P.1366 - 1371
5. **Lladoós, J.** Symbol recognition by error-tolerant subgraph matching between region adjacency graphs / J. Lladoós, E. Martí, J. Villanueva //IEEE Transactions on Pattern Analysis and Machine Intelligence, 2001. - Vol.23. - No.10. - P.1137-1143.
6. **Ryu, Y.S.** Automatic extraction of eye and mouth fields from a face image using eigenfeatures and multilayer perceptrons / Y.S. Ryu, S.Y. Oh //Pattern recognition, 2001. - Vol.34. - No. 12. - P. 2459-2466.
7. **Su Ya** Symbol recognition via statistical integration of pixel-level constraint histograms: a new descriptor /Su Ya//IEEE Transactions on Pattern Analysis and Machine Intelligence archive, 2005. - Vol.27. - Issue 2. - P. 278 - 281.

METHOD OF RAPID CORRELATION USING TERNARY TEMPLATES FOR OBJECT RECOGNITION IN IMAGES

N.I. Glumov¹, E.V. Myasnikov^{1,2}, V.N. Kopenkov^{1,2}, M.A. Chicheva¹

¹Image Processing Systems Institute of RAS, Samara, Russia,

²Samara state aerospace university, Samara, Russia

Abstract

The task of recognition and position detection of fragments corresponding to one of templates of a set is considered in this work. The proposed method of fast correlation using a set of ternary templates solve the problems indicated above successfully. The method is applied to solve the two following tasks: the first task is detection of eye position on the face facsimile on the documents and the second one is recognition of machine readable text on the scanned documents.

Key words: correlation, object recognition, document image, eye search, machine-readable data.

Citation: Glumov NI, Myasnikov EV, Kopenkov VN, Chicheva MA. Method of rapid correlation using ternary templates for object recognition in images. Computer Optics 2008; 32(3): 277-82.

Acknowledgements: The work was supported by the Russian Foundation for Basic Research (RFBR), projects № 06-01-00722, 06-01-00616, 07-01-96612, 07-07-97610, 08-07-90704-mob_st within the framework of Russian-American program "Basic Research and Higher education» (CRDF Project RUX0-014-SA-06) and a grant from the Russian President on the sub-support of leading scientific schools (NSH-3086.2008.9).

References

- [1] Document ICAO 9303 "Machine Readable Travel Documents (MRTD)" //http://www.icao.int/mrtd/publications/doc.cfm
- [2] Soifer VA, ed. Methods of computer image processing [In Russian]. Moscow: "Fizmatlit" Publisher 2001; 784 p.
- [3] Chicheva MA, Glumov NI, Sergeev VV. Optimization of linear filtering procedure in image processing applications. Proceedings of the 6-th German-Russian workshop "Pattern recognition and image understanding" OGRW-6-2003; 2003: 149-154.
- [4] Kawato S, Ohya J. Two-step approach for real-time eye tracking with a new filtering technique. Systems, Man and Cybernetics 2000; 2: 1366-1371.
- [5] Lladoós J, Martí E, Villanueva J. Symbol recognition by error-tolerant subgraph matching between region adjacency graphs. IEEE Transactions on Pattern Analysis and Machine Intelligence 2001; 23(10): 1137-1143.
- [6] Ryu YS, Oh SY. Automatic extraction of eye and mouth fields from a face image using eigenfeatures and multilayer perceptrons. Pattern recognition 2001; 34(12): 2459-2466.
- [7] Su Ya. Symbol recognition via statistical integration of pixellevel constraint histograms: a new descriptor. IEEE Transactions on Pattern Analysis and Machine Intelligence archive 2005; 27(2): 278 - 281.