

Алгоритм идентификации объектов по изображениям, основанный на разделении гиперплоскостью и нечувствительный к изменению условий освещения и ракурса

А. В. Зубюк^а, А. Б. Федотов^б

Московский государственный университет имени М. В. Ломоносова, физический факультет, кафедра математического моделирования и информатики.

Россия, 119991, Москва, Ленинские горы, д. 1, стр. 2.

E-mail: ^а zubuk@cmpd2.phys.msu.ru, ^б toivo@ya.ru

Статья поступила 13.07.2015, подписана в печать 21.08.2015.

Предложен алгоритм идентификации объектов по их изображениям, сочетающий сильные стороны методов морфологического анализа изображений Пытьева, а именно инвариантность относительно изменения неконтролируемых условий регистрации изображений, с вычислительной эффективностью и простотой методов, основанных на разделении классов изображений гиперплоскостями. Работа алгоритма продемонстрирована на изображениях объектов, снятых при разных условиях освещения и ракурсах. Алгоритм может быть применен для интерпретации микроскопических изображений в биофизике, анализа формы акустических сигналов в геофизике, исследования природных ресурсов Земли по спутниковым снимкам и т. д.

Ключевые слова: морфологический анализ изображений, распознавание образов, разделяющая гиперплоскость, метод опорных векторов.

УДК: 004.93. PACS: 07.05.Pj, 07.05.Mh.

Введение

Трудности при решении многих задач анализа изображений и сигналов связаны с тем, что зарегистрированные изображения и сигналы, помимо полезной для исследователя информации, несут также информацию об условиях их регистрации, которые исследователю не известны и могут неконтролируемо изменяться. Так, например, изображение какого-либо объекта несет информацию не только о его геометрической форме, ориентации в пространстве и других характеристиках, присущих непосредственно этому объекту, но также и об условиях освещения, при которых получено данное изображение, характеристиках и настройках использованной камеры и т. п. Пусть решается задача опознавания изображенного объекта, тогда условия освещения и характеристики камеры следует считать «мешающими» параметрами, изменение которых не должно влиять на результат узнавания. Приведем еще один пример. Порожденный удаленным источником — таким как ядерный взрыв или горный обвал — акустический сигнал, распространяясь в атмосфере, претерпевает значительные искажения, вследствие чего несет информацию не только об источнике звука, но и об акустических свойствах атмосферы [1]. При решении задачи идентификации источника звука «мешающими» параметрами следует считать акустические свойства атмосферы, и их изменение не должно влиять на результат идентификации.

Для решения подобных задач были разработаны методы морфологического анализа изображений и сигналов [2–8], в рамках которых изображения и сигналы различной природы рассматриваются как

элементы функциональных пространств, что позволяет использовать для их описания и анализа универсальные методы линейной алгебры и функционального анализа. В связи с этим далее в настоящей работе будем говорить об анализе изображений, подразумевая, что те же методы могут быть применены и для анализа различных сигналов. В основе методов морфологического анализа лежит идея инвариантности относительно изменения неизвестных и неконтролируемых условий регистрации. Максимальный инвариант класса всех таких изменений называется *формой изображения*, она может быть определена как множество всех изображений, несущих одну и ту же полезную информацию и отличающихся лишь условиями их регистрации. При решении прикладных задач методами морфологического анализа используется техника ортогонального или косоугольного проецирования на форму изображения, что определяет их вычислительную эффективность в случаях, когда форма изображения представляет собой достаточно простое множество, например, линейное подпространство малой размерности. Однако если форма изображения не является линейным подпространством и имеет большую размерность, задача проецирования на нее становится достаточно сложной как в математическом, так и в вычислительном плане.

Другим распространенным подходом к решению некоторых задач интерпретации изображений и сигналов, таких как идентификация объектов по их изображениям или идентификация типов источников звука по зарегистрированным акустическим сигналам, является разделение классов изображений

или сигналов достаточно простыми многообразиями (например, гиперплоскостями) в пространстве признаков [9–14]. Данный подход прост в реализации и позволяет строить более эффективные с вычислительной точки зрения алгоритмы по сравнению с алгоритмами, основанными на свойственной методам морфологического анализа технике проецирования. Основная трудность данного подхода состоит в отыскании признаков, т.е. числовых характеристик изображения, использование которых позволит произвести желаемое разделение классов изображений. Известны методы, позволяющие произвести отбор подобных признаков автоматически [15–19]. Однако в большинстве задач анализа изображений удачный выбор признаков зависит от интуиции и опыта исследователя [20–27]. Кроме того, такие признаки зачастую оказываются неинвариантными относительно всевозможных изменений условий регистрации.

В настоящей работе рассмотрена модель регистрации изображений объектов, освещаемых некогерентными источниками света; при этом предполагается, что каждый светочувствительный детектор используемой для регистрации камеры дает сигнал на выходе, пропорциональный падающему на него световому потоку. В результате исследования данной модели было показано, что множества изображений разных объектов, полученных при всевозможных условиях освещения и ракурсах, могут быть разделены гиперплоскостями непосредственно в пространстве изображений. Это позволило разработать алгоритм, основанный на разделении гиперплоскостью и нечувствительный к изменению условий освещения и ракурса, позволяющий идентифицировать объекты по их изображениям и автоматически обучаться идентификации. В работе приведены результаты тестирования разработанного алгоритма на реальных изображениях и сравнения его качества с широко известным методом опорных векторов [13, 14].

1. Модель регистрации изображений

Полем зрения будем называть конечное подмножество $\mathcal{X} = \{x_1, \dots, x_n\}$ координатной плоскости $(-\infty, \infty) \times (-\infty, \infty)$, состоящее из точек $x_1, \dots, x_n$, называемых пикселями. *Изображением* будем называть всякую функцию $f: \mathcal{X} \rightarrow (-\infty, \infty)$, значения $f(x)$ которой будем называть *яркостями изображения* f , $x \in \mathcal{X}$. Линейную комбинацию изображений f и g с коэффициентами α и β соответственно определим как изображение

$$x \mapsto \alpha f(x) + \beta g(x),$$

а их скалярное произведение (f, g) — как число

$$(f, g) = \sum_{x \in \mathcal{X}} f(x)g(x).$$

Множество всех изображений с определенными выше линейными операциями и скалярным произведе-

нием образует евклидово пространство размерности n , которое будем называть *пространством изображений* и обозначать $\mathcal{R}$.

Рассматриваемая в настоящей работе модель регистрации изображений определяется следующими условиями, налагаемыми на используемую для регистрации камеру, оптические свойства снимаемых объектов, а также источники света.

1. *Используемая камера.* Яркость $f(x)$ зарегистрированного камерой изображения f в точке $x \in \mathcal{X}$ пропорциональна световому потоку, падающему на светочувствительный детектор, соответствующий пикселю x .

2. *Оптические свойства снимаемых объектов.* Снимаемые объекты не излучают свет сами по себе, а лишь отражают падающий на них свет. При освещении объекта точечным источником света для каждой точки поверхности объекта справедлива следующая связь между интенсивностями падающего и отраженного света:

$$I_{\text{ref}}(\mathbf{e}) = k I_{\text{inc}}(\mathbf{s}), \quad (1)$$

где $\mathbf{s}$ и $\mathbf{e}$ — единичные векторы, определяющие направления падения и отражения света соответственно; $I_{\text{inc}}(\mathbf{s})$ и $I_{\text{ref}}(\mathbf{e})$ — интенсивности света, падающего в направлении $\mathbf{s}$ и отраженного в направлении $\mathbf{e}$ соответственно; k — коэффициент отражения света, не зависящий от $I_{\text{inc}}(\mathbf{s})$.

3. *Источники света.* Освещение, при котором проводится регистрация, формируется точечными некогерентными источниками света.

2. Форма изображения объекта, регистрируемого при неизменном ракурсе

Рассмотрим регистрацию изображений одного и того же объекта при неизменном ракурсе съемки. Совокупность скалярных параметров, определяющих выбранный ракурс, обозначим φ . Такими параметрами могут быть, например, два угла, задающих ориентацию снимаемого объекта в пространстве. Варьироваться могут лишь условия освещения объекта. Сформулированные выше ограничения 1–3 позволяют определить, что представляет собой множество всех изображений объекта, полученных при всевозможных условиях освещения, являющееся в данном случае *формой изображения* объекта [6].

Действительно, пусть изображение $f \in \mathcal{R}$ рассматриваемого объекта получено при условиях освещения L . Изменим в $\alpha \geq 0$ раз яркости всех источников света, формирующих освещение L , и обозначим новые условия освещения αL . При изменении освещения L на αL изменяется в α раз интенсивность падающего на объект света, следовательно, в силу (1), в α раз изменяется и интенсивность отраженного света, а значит и световой поток, падающий на каждый детектор камеры. Поэтому согласно условию 1, налагаемому на камеру, изображение объекта, полученное при освещении αL , есть αf .

Таким образом, если f — изображение объекта, то αf — изображение того же объекта, полученное, возможно, при других условиях освещения, $\alpha \geq 0$.

Рассмотрим теперь изображения $f_1 \in \mathcal{R}$ и $f_2 \in \mathcal{R}$ рассматриваемого объекта, полученные при условиях освещения L_1 и L_2 (наличие какой-либо связи между освещенностями L_1 и L_2 в данном случае не предполагается). Пусть L_{1+2} есть освещение, представляющее собой суперпозицию освещений L_1 и L_2 , т.е. L_{1+2} получено одновременным включением источников света, формирующих освещение L_1 и L_2 . В силу линейности (1) и некогерентности источников света (условие 3) световой поток, падающий на каждый детектор камеры при освещении L_{1+2} есть сумма световых потоков, формируемых освещенностями L_1 и L_2 в отдельности. Поэтому изображение объекта, полученное при освещении L_{1+2} , есть $f_1 + f_2$. Таким образом, если f_1 и f_2 — изображения одного объекта, то $f_1 + f_2$ — изображение того же объекта.

Суммируя вышесказанное, получаем, что форма изображения объекта, снимаемого при неизменном ракурсе φ , есть множество $V_\varphi \subset \mathcal{R}$, удовлетворяющее условиям

$$\begin{aligned} f \in V_\varphi &\Rightarrow \alpha f \in V_\varphi, \quad \alpha \geq 0, \\ f_1, f_2 \in V_\varphi &\Rightarrow f_1 + f_2 \in V_\varphi. \end{aligned}$$

То есть форма изображения V_φ является выпуклым конусом [28].

3. Форма изображения объекта, регистрируемого при изменяющемся ракурсе

Пусть теперь изменяться могут не только условия освещения объекта, но и ракурс съемки φ . Множество, в пределах которого может принять свое значение векторный параметр φ , обозначим Φ . Форма изображения V рассматриваемого объекта в этом случае может быть выражена следующим образом:

$$V = \bigcup_{\varphi \in \Phi} V_\varphi,$$

где V_φ — выпуклый конус, представляющий собой множество изображений объекта, зарегистрированных при фиксированном ракурсе $\varphi \in \Phi$ и всевозможных условиях освещения. Такое множество V является конусом, но, вообще говоря, не выпуклым (закрашенная серым область на рис. 1).

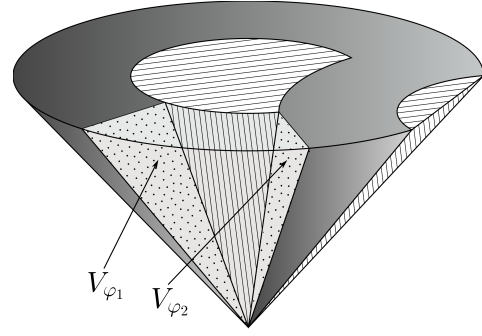


Рис. 1. Наглядное представление формы изображения объекта, снимаемого при всевозможных условиях освещения и ракурсах. Точками выделены выпуклые конусы V_{φ_1} и V_{φ_2} , каждый из которых есть множество изображений объекта, зарегистрированных при одном ракурсе, но разных условиях освещения. Форма изображения объекта (выделена серым) есть объединение выпуклых конусов, соответствующих всевозможным ракурсам съемки; она представляет собой конус, но, вообще говоря, не выпуклый. Ее дополнение до выпуклого конуса (выделено штриховкой) состоит из изображений, которые не могут быть зарегистрированы в реальности

Обозначим $\bar{V}$ выпуклую оболочку формы V , которая в силу теоремы Каратеодори [28] и того, что размерность $\dim \mathcal{R} = n$ пространства изображений $\mathcal{R}$ конечна, может быть представлена в виде

$$\bar{V} = \left\{ \sum_{i=1}^{n+1} \alpha_i f_i \mid \alpha_i \geq 0, f_i \in V, i = 1, \dots, n+1, \sum_{i=1}^{n+1} \alpha_i = 1 \right\} \quad (2)$$

и в рассматриваемом случае является выпуклым конусом. Обозначим δV дополнение формы V до ее выпуклой оболочки $\bar{V}$, т.е. $\delta V = \bar{V} \setminus V$. На рис. 1 множество δV обозначено штриховкой. В соответствии с (2) всякое изображение f из δV может быть представлено в виде выпуклой комбинации некоторых изображений рассматриваемого объекта, среди которых найдутся изображения, зарегистрированные при разных ракурсах и входящие в комбинацию с положительными коэффициентами (иначе $f \in V_\varphi \subset V$ при некотором $\varphi \in \Phi$). Пример такого изображения для случая, когда снимаемым объектом является гипсовая фигурка медвежонка, приведен на рис. 2. Как видно из рисунка, подобные изображе-



Рис. 2. Пример нереалистичного изображения, являющегося суммой двух изображений одного объекта, зарегистрированных при разных ракурсах

ния выглядят как наложения принципиально разных изображений, в связи с чем далее будем считать, что они не являются изображениями каких-либо реальных объектов и не могут быть зарегистрированы.

Суммируя вышесказанное, получаем, что форма изображения объекта, снимаемого при произвольном ракурсе, есть конус V , выпуклая оболочка которого $\bar{V}$ не содержит никаких реальных изображений, кроме изображений рассматриваемого объекта.

4. Идентификация и обучение идентификации

Пусть имеются два физических объекта O_1 и O_2 , изображения которых, снятые при любых ракурсах и любых условиях освещения, кроме полного отсутствия освещения, отличаются друг от друга. Рассмотрим задачу идентификации данных объектов по их изображениям. Итак, пусть предъявлено изображение $g \in \mathcal{R}$ одного из объектов O_1, O_2 , требуется определить, изображением какого из двух объектов оно является.

Обозначим $V^{(i)}$ форму изображения объекта O_i , регистрируемого при изменяющихся условиях освещения и ракурсе, $i = 1, 2$. В соответствии с выводами предыдущего раздела выпуклая оболочка $\bar{V}^{(i)}$ формы $V^{(i)}$, являющаяся выпуклым конусом, содержит реальные изображения только объекта O_i и не содержит изображений объекта O_j , т.е. $\bar{V}^{(i)} \cap V^{(j)} = \{0\}$, $j \neq i$. Сделаем более сильное, но вполне реалистичное предположение, состоящее в том, что выпуклая комбинация любых изображений объекта O_1 не может быть представлена в виде выпуклой комбинации изображений объекта O_2 и наоборот, т.е. что $\bar{V}^{(i)} \cap \bar{V}^{(j)} = \{0\}$, $j \neq i$. Учитывая все сделанные предположения и все перечисленные свойства множеств $\bar{V}^{(i)}$, получаем, что $\bar{V}^{(1)}$ и $\bar{V}^{(2)}$ могут быть разделены гиперплоскостью в пространстве изображений $\mathcal{R}$, проходящей через начало координат. То есть найдется такое изображение $w \in \mathcal{R}$, $\|w\| > 0$, что

$$\begin{aligned} (g, w) &\leq 0 \quad \text{при } g \in \bar{V}^{(1)}, \\ (g, w) &\geq 0 \quad \text{при } g \in \bar{V}^{(2)}, \end{aligned} \quad (3)$$

причем $(g, w) = 0$, если и только если $g \in \bar{V}^{(1)} \cap \bar{V}^{(2)} = \{0\}$.

Поскольку $V^{(i)} \subset \bar{V}^{(i)}$ и изображение $g = 0$ соответствует исключенной из рассмотрения ситуации полного отсутствия освещения, неравенства (3) могут быть использованы для решения задачи идентификации следующим образом:

$$\begin{aligned} \text{если } (g, w) < 0, \quad &\text{то } g \text{ — изображение } O_1, \\ \text{если } (g, w) > 0, \quad &\text{то } g \text{ — изображение } O_2. \end{aligned} \quad (4)$$

Для поиска изображения $w \in \mathcal{R}$, $\|w\| > 0$, в (4) может быть использован следующий эвристический алгоритм. Пусть

$$f_1^{(i)}, \dots, f_{L_i}^{(i)} \quad (5)$$

— изображения объекта O_i , полученные при разных условиях освещения и ракурсах, $i = 1, 2$. Множество изображений (5) будем называть *обучающей выборкой*. Определим w следующим образом:

$$w = e^{(2)} - e^{(1)}, \quad \text{где } e^{(i)} = \frac{1}{L_i} \sum_{j=1}^{L_i} f_j^{(i)} / \|f_j^{(i)}\|, \quad i = 1, 2. \quad (6)$$

Такой выбор w основан на следующих соображениях. Если условия освещения и ракурсы, при которых зарегистрированы изображения из обучающей выборки, таковы, что $f_j^{(i)} / \|f_j^{(i)}\|$, $j = 1, \dots, L_i$, «равномерно разбросаны» по пересечению конуса $V^{(i)}$ с единичной сферой, то вектор $e^{(i)}$ расположен на луче, лежащем «в центре» конуса $V^{(i)}$. В связи с этим можно рассчитывать, что гиперплоскость, ортогональная изображению w , соединяющему¹ $e^{(1)}$ и $e^{(2)}$, разделит конусы $V^{(1)}$ и $V^{(2)}$.

5. Тестирование алгоритма идентификации и обучения идентификации

Для тестирования предложенного выше алгоритма было использовано 240 изображений гипсовых фигурок, аналогичных приведенным на рис. 2 слева и в центре, из них 120 — изображения гипсовой фигурки медвежонка (объект O_1), и еще 120 — изображения гипсовой фигурки девушки (объект O_2). Изображения обеих фигурок были зарегистрированы при разных условиях освещения и ракурсах.

Имеющееся множество изображений случайным образом разбивалось на две части.

1. *Обучающая выборка.* Для обучения, т.е. для построения изображения w по формуле (6), отбирались $L = L_1 = L_2$ изображений каждой фигурки.

2. *Тестовая последовательность.* Оставшиеся $240 - 2L$ изображений использовались для тестирования алгоритма идентификации, работающего согласно (4).

Такое случайное разбиение производилось многократно. По результатам проведенных тестов рассчитывалась средняя частота α ошибок идентификации, т.е. отношение количества изображений из тестовой последовательности, для которых фигурка медвежонка была опознана как фигурка девушки или наоборот, к общему количеству изображений, использованных для тестирования. Зависимость α от размера L обучающей выборки приведена на рис. 3 (сплошная линия).

Кроме этого, на рис. 3 штрихпунктирной линией отображена аналогичная зависимость для другого широко применяемого алгоритма идентификации

¹ Слово «соединять» здесь применено в связи с геометрической интерпретацией векторной разности в трехмерном пространстве, в соответствии с которой вектор $c = b - a$ соединяет концы векторов a и b .

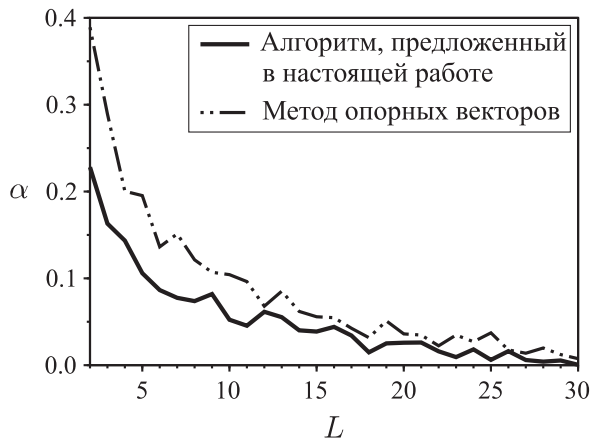


Рис. 3. Зависимость средней частоты α ошибок идентификации от размера L обучающей выборки для алгоритма, предложенного в настоящей работе (сплошная линия), и метода опорных векторов (штрих-пунктирная линия)

и обучения идентификации, использующего разделение классов изображений гиперплоскостью, — метода опорных векторов (support vector machine, SVM) [13, 14]. В отличие от алгоритма, предложенного в настоящей работе, SVM разделяет классы изображений гиперплоскостью, не проходящей, вообще говоря, через начало координат, и не использует при ее построении каких-либо представлений о модели регистрации изображений.

Как видно из рис. 3, алгоритм, предложенный в настоящей работе, демонстрирует меньшую среднюю частоту ошибок, чем метод опорных векторов.

Заключение

Разработан алгоритм идентификации и обучения идентификации объектов по их изображениям, зарегистрированным при разных условиях освещения и ракурсах. Работоспособность алгоритма обоснована как теоретически путем анализа модели регистрации изображений, так и практически в ходе тестирования на реальных изображениях, при котором предложенный алгоритм продемонстрировал более высокое качество идентификации в сравнении с методом опорных векторов.

Дальнейшее развитие алгоритма может состоять в его адаптации к практически важному случаю, когда неизвестными и неконтролируемыми являются не только условия освещения и ракурс съемки объектов, но и характеристики светочувствительных детекторов используемой для съемки камеры.

Работа выполнена при финансовой поддержке РФФИ, грант № 14-07-00409 А.

Список литературы

1. Куличков С.Н., Чуликов А.И., Демин Д.С. Морфологический анализ инфразвуковых сигналов в акустике. М., 2010.
2. Пытьев Ю.П. // ДАН СССР. 1975. **224**, № 6. С. 1283.
3. Пытьев Ю.П. // ДАН СССР. 1983. **269**, № 5. С. 1061.
4. Pyt'ev Yu.P. // Pattern Recognition and Image Analysis. 1993. **3**, N 1. P. 19.
5. Pyt'ev Yu.P. // Pattern Recognition and Image Analysis. 1998. **8**, N 4. P. 517.
6. Пытьев Ю.П., Чуликов А.И. Методы морфологического анализа изображений. М., 2010.
7. Визильтер Ю.В., Желтов С.Ю. // Изв. РАН. ТиСУ. 2008. № 6. С. 113.
8. Vizilter Yu.V., Zheltov S.Yu. // J. Computer and Systems Sciences International. 2009. **48**, N 2. P. 282.
9. Ту Дж., Гонсалес Р. Принципы распознавания образов. М., 1978.
10. Вапник В.Н., Червоненкис А.Я. Теория распознавания образов. М., 1974.
11. Bishop C.M. Pattern Recognition and Machine Learning. Springer Science & Business Media, 2006.
12. Fisher R.A. // Annals of Eugenics. 1936. **7**, N 2. P. 179.
13. Boser B.E., Guyon I., Vapnik V.N. // Proc. of the 5-th Annual Workshop of Computational Learning Theory. 1992. **5**. P. 144.
14. Cortes C., Vapnik V. // Machine learning. 1995. **20**, N 3. P. 273.
15. Pearson K. // The London, Edinburgh, and Dublin Philosophical Magazine and J. of Sci. 1901. **6**, N 2. P. 559.
16. Hotelling H. // J. of Educational Psychology. 1933. **24**, N 6. P. 417.
17. Jolliffe I. Principal component analysis. Springer, 2002.
18. Tenenbaum J.B., De Silva V., Langford J.C. // Science. 2000. **290**, N 5500. P. 2319.
19. Lee J.A., Verleysen M. Nonlinear Dimensionality Reduction. Springer Science & Business Media, 2007.
20. Chellappa R., Bagdazian R. // IEEE Trans. Pattern Analysis and Machine Intelligence. 1984. **6**, N 1. P. 102.
21. Zhang G., Ma Z.M., Tong Q. et al. // Proc. of Int. Conf. on Intelligent Information Hiding and Multimedia Signal Processing. 2008. P. 71.
22. Boles W.W. // Fourth Internat. Symposium on Signal Processing and Its Applications. 1996. **2**. P. 601.
23. Szewczyk R., Grabowski K., Napieralska M. et al. // Pattern Recognition Lett. 2012. **33**, N 8. P. 1019.
24. Abbasi S., Mokhtarian F., Kittler J. // Multimedia Systems. 1999. **7**, N 6. P. 467.
25. Daoudi M., Matusiak S. // J. of Visual Languages & Computing. 2000. **11**, N 3. P. 287.
26. Местецкий Л.М. // Машинное обучение и анализ данных. 2014. **1**, № 9. С. 1291.
27. Сидякин С.В., Визильтер Ю.В. // Компьютерная оптика. 2014. **38**, № 3. С. 511.
28. Рокафеллар Р. Выпуклый анализ. М., 1973.

An algorithm for object identification based on hyperplane separation but tolerant to illumination and viewing angle variations**A.V. Zubuk^a, A.B. Fedotov^b***Department of Mathematical Modeling and Informatics, Faculty of Physics, Lomonosov Moscow State University, Moscow 119991, Russia.**E-mail: ^a zubuk@cmpd2.phys.msu.ru, ^b toivo@ya.ru.*

A new algorithm for object identification by their images is proposed. It combines tolerance against imaging conditions variations proper to morphological methods for image analysis by Pyt'yev with computational simplicity of methods based on separating classes of images with hyperplane. The algorithm was tested on images taken in different illuminations and views. It can be applied to analyse microscopic image in biophysics, acoustic signals in geophysics, Earth resources using satellite data, etc.

Keywords: morphological image analysis, pattern recognition, hyperplane separation, support vector machine.

PACS: 07.05.Pj, 07.05.Mh.

Received 13 June 2015.

English version: *Moscow University Physics Bulletin* 6(2015).

Сведения об авторах

1. Зубюк Андрей Владимирович — мл. науч. сотрудник; тел.: (495) 939-46-94, e-mail: zubuk@cmpd2.phys.msu.ru.
2. Федотов Андрей Борисович — студент; e-mail: toivo@ya.ru.