УДК 004.932

ВОССТАНОВЛЕНИЕ СИЛУЭТА РУКИ В ЗАДАЧЕ РАСПОЗНАВАНИЯ ЖЕСТОВ С ПОМОЩЬЮ АДАПТИВНОЙ МОРФОЛОГИЧЕСКОЙ ФИЛЬТРАЦИИ БИНАРНОГО ИЗОБРАЖЕНИЯ

© 2013 г. **Р. О. Малашин**, аспирант; **В. Р. Луцив**, доктор техн. наук

ОАО "Государственный оптический институт им. С. И. Вавилова", Санкт-Петербург

E-mail: malashinroman@mail.ru

Приведены алгоритмы адаптивной обработки бинарных изображений силуэтов рук человека, полученных с помощью цвето-яркостных фильтров. Приведенные алгоритмы основаны на использовании комбинации элементарных морфологических операций с учетом направления пальцев руки. Приведены алгоритмы удаления шума на бинарных изображениях, адаптирующихся к результату работы цветового фильтра, и способ заполнения внутренних контуров силуэта руки с целью удаления групповых ошибок маркирования. Результаты экспериментов показывают, что предложенный способ обработки изображений повышает вероятность успешного обнаружения, слежения за рукой и распознавания жестов.

Ключевые слова: *распознавание жестов руки, восстановление силуэта руки, морфологическая обработка бинарных изображений.*

Коды OCIS: 120.3930, 260.7210, 300.6210, 300.6540.

Поступила в редакцию 17.07.2013.

Введение

Жесты рук являются важной частью невербального общения людей и поэтому автоматический анализ жестов может быть полезен при построении человеко-машинных интерфейсов и для проведения психологических исследований [1]. Особое место занимают технологии, использующие видео- и фотокамеры, поскольку они удобны для пользователя. Для создания таких технологий необходимо обнаруживать и следить за рукой человека в кадре. Цвет кожи компактно располагается в различных цветовых пространствах и хорошо подходит для выделения руки в видеопоследовательности, поскольку методы, основанные на цвете пикселов изображения, инвариантны к преобразованиям формы ладони и пальцев. Результатом работы цветового фильтра является сегментированное изображение, на котором силуэт руки и фон выделены различными метками. Сегментированные изображения могут быть проанализированы и использованы для распознавания статических [2] и динамических жестов [3]. Для повышения надежности слежения можно использовать трехмерную модель внешнего вида руки. При таком подходе из видеопоследовательности извлекают силуэт руки, а затем итеративно минимизируют меру несоответствия полученного силуэта с силуэтом, полученным с помощью трехмерной модели, путем подстройки ее параметров [4, 5].

При выделении изображения искомого объекта из фона на основе сегментации изображений по цвето-яркостным или текстурным признакам возникают многочисленные локальные ошибки обнаружения и маркирования областей изображения в виде шума типа "перец и соль" в областях выделенного объекта и фона, к нему не относящегося, соответственно (здесь и далее полагается, что на бинарном изображении значение "1" соответствует меткам руки, а "0" – меткам фона). Причина этого заключается в отличии освещения от эталонного и в сложных фонах. Кроме того, цвет кожи ладони рук различен у разных людей, а каждая ладонь имеет участки разного цвета. Именно поэтому силуэты, полученные с помощью цветовых фильтров, имеют как одиночные, так и групповые ошибки маркирования, которые в конечном итоге могут привести к неправильному обнару-

жению руки или распознаванию жеста. Шумовые дефекты сегментации обычно удаляются методами морфологической фильтрации [6]. При этом на итоговый результат ключевое влияние оказывает размер структурирующего элемента, применяемого для фильтрации. Чем он больше, тем более существенный шум может быть удален (а групповые ошибки маркирования исправлены) в результате морфологической фильтрации. В то же время с увеличением структурирующего элемента все больше деградирует форма силуэта объекта, что негативно сказывается при его распознавании и в других задачах. При выборе универсального размера структурирующего элемента морфологическая фильтрация либо недостаточно хорошо подавляет шум на сильно зашумленных изображениях (что делает невозможным дальнейший анализ изображений), либо повреждает силуэты объектов на слабо зашумленных изображениях. В данной работе эта проблема решается путем построения морфологического фильтра с адаптивным выбором размера и формы структурирующего элемента. Были получены хорошие результаты восстановления силуэтов рук по базе данных более чем из 450 изображений.

Получение силуэтов руки по цветному изображению

В данной работе цветовая сегментация изображения ладони руки была основана на работе двух цветовых фильтров. Первый из них использовал цветовое пространство rg [7, 8], второй – цветовые интервалы в пространстве HSV [9]. В результате объединения результатов работы двух цветовых фильтров (автоматического выбора лучшего из них) были получены бинарные маски, на которых были отмечены пикселы цвета кожи. Из-за разных условий съемки и применяемых цветовых фильтров полученные маски сильно отличались друг от друга: некоторые из них были получены с очень большой точностью (рис. 1а) – шумовые эффекты практически отсутствовали, другие, наоборот, были подвержены сильному шуму (рис. 1в). Для второго типа масок дополнительная обработка была необходима, в то время как для первого типа требовалось лишь удаление незначительного шума. Для решения этой задачи была разработана процедура адаптивной морфологической фильтрации.

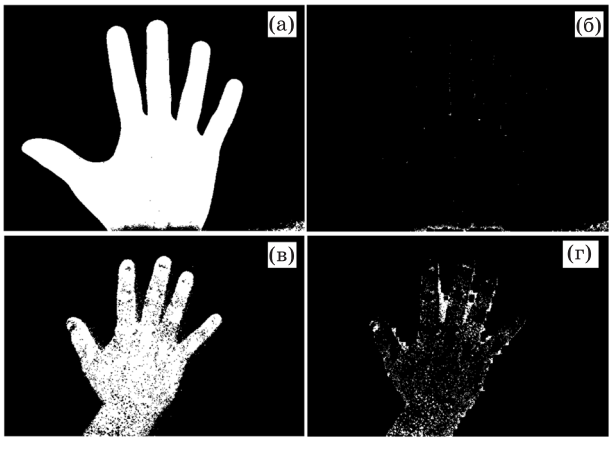

Рис. 1. Исходные изображения – а, в, б, г – результат операции CLOSE за исключением элементов исходного изображения.

Морфологическая фильтрация

Известно, что морфологические операции OPEN и CLOSE могут удалять шумы на бинарных изображениях. Но, поскольку нет никакой априорной информации об уровне шума, сложность представляется в определении правильного размера структурирующего элемента этих изображений. Кроме того, разные типы шумов требуют разных комбинаций морфологических операций. Можно выделить два типа ошибок (шумов), сопровождающие сегментированные изображения

1. Шум типа “перец” и “соль” – одиночные ошибочные решения о принадлежности пиксела к изображению ладони руки.

2. “Заливы” и “острова” – групповые ошибочные решения о принадлежности пиксела к изображению ладони руки.

Таким образом, морфологическая фильтрация должна решать следующие задачи:

1. Удалять шум и при этом не повреждать границы силуэтов, которые были обнаружены цветовым фильтром правильно.

2. Восстанавливать плохо сегментированные изображения, чтобы была возможность распознать жест (в этом случае допустимы незначительные смещения границ силуэта относительно истинных границ руки).

Удаление шума

Для измерения уровня зашумленности исходного бинарного изображения применяется операция CLOSE, а затем из полученного изображения исключаются все пикселы, которые присутствовали на исходном изображении

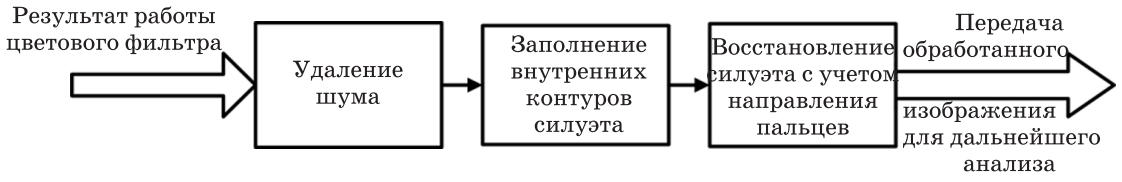

Рис. 2. Общая схема предлагаемого алгоритма обработки бинарных изображений.

$$D = \text{CLOSE}(S) - S,$$

где S – исходное бинарное изображение, D – результирующее бинарное изображение.

Таким образом, на бинарном изображении остаются только те пикселы, которые добавляются в результате операции CLOSE. Чем сильней исходная маска повреждена случайными выбросами, тем большее количество новых пикселов будет порождено в результате описанных выше операций (рис. 1). В соответствии с этим рассчитывается коэффициент зашумленности изображения

$$C_n = W_D/W_S,$$

где C_n – коэффициент зашумленности изображения, W_S – количество меток руки на исходном изображении, W_D – количество меток руки после обработки. Значение C_n используется для выбора размера структурирующего элемента, используемого для восстановления изображения, а также для выбора процедуры удаления шума. Общая схема предлагаемого алгоритма обработки бинарных изображений представлена на рис. 2.

В настоящей работе улучшение результатов морфологической фильтрации достигается с помощью построения двух фильтров для подавления шума и адаптивного выбора одного из них в зависимости от коэффициента C_n и от результатов их работы.

Фильтр “А” является морфологической операцией условного наращивания (*conditional dilation*) и состоит из следующих примитивных операций:

1. ERODE (a[1]),
2. DILATE (b),
3. Бинарная операция AND для получившегося и исходного изображений.

Фильтр “Б” представляет собой последовательность примитивных морфологических операций

1. DILATE (a),
2. ERODE (b),
3. DILATE (b–a).

Первый фильтр (“А”) использует операцию условного наращивания и предназначен для изображений, где область объекта не подвержена значительным шумам. Операция условного наращивания состоит из последовательности примитивных бинарных и морфологических операций. Сначала выполняется операция ERODE, которая позволяет удалить шум в области фона. После этого выполняется операция DILATE с бóльшим структурирующим элементом, что позволяет увеличить область объекта и частично восстановить достоверные, но удаленные в результате эрозии метки объекта. Чтобы исключить появление меток объекта, не присутствовавших на исходном сегментированном изображении, для полученного бинарного изображения выполняется бинарная операция AND с исходным изображением. В результате обеспечивается более аккуратное удаление шумов на фоне изображения, чем при использовании операции OPEN (рис. 3).

Второй фильтр (“Б”) использует последовательность примитивных морфологических операций и обеспечивает еще более “мягкое” удаление меток объекта на фоне. Для этого сначала выполняется операция DILATE, которая позволяет объединить одиночные метки объекта, находящиеся близко друг к другу. После этого выполняется операция ERODE со структурирующим элементом бóльшего размера. Таким образом, удаляются только те метки объекта, которые на исходном изображении находились далеко от других меток объекта, а метки объекта, располагавшиеся близко к другим меткам

[1] Здесь и далее в скобках указан размер квадратного структурирующего элемента, причем $b > a$.

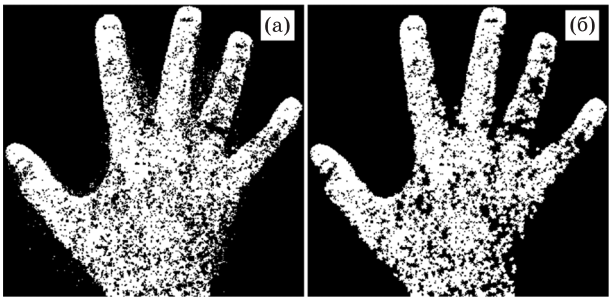

Рис. 3. Исходное изображение – а, б – изображение после удаления шума с помощью операции условного наращивания.

объекта (пусть даже не связанные с ними непосредственно) сохраняются. Для восстановления исходного размера области объекта на последнем шаге выполняется операция ERODE со структурирующим элементом соответствующего размера.

Размер структурирующих элементов зависит от размеров силуэта руки (количества меток руки). В настоящей работе для изображений размером 640×480, где проекция кисти руки имеет сопоставимые размеры, были использованы размеры структурирующих элементов $a = 3$ и $b = 5$.

Выбор конкретного фильтра осуществляется по коэффициенту C_n: если он меньше заданного порога, к изображению применяется фильтр “А”, а если больше – фильтр “Б”. Кроме того, если в результате выполнения фильтра “А” удалено больше 15% меток руки, то изменения отменяются и применяется фильтр “Б”. Аналогично, если фильтр “Б” удаляет больше 15% меток руки, на дальнейшую обработку подается исходное бинарное изображение без удаления шума.

Удаление групповых ошибок маркирования

В связи с тем, что кожа человека на одной руке может иметь разные оттенки, очень часто на сегментированном изображении в области соответствующей ладони, могут появляться скопления меток фона. Такие области, как правило, соответствуют определенному оттенку кожи, характерному для данного участка руки (или блику на руке), и ошибочно интерпретируются цветовым фильтром как фон. Такие шумовые дефекты сегментации могут отрицательным образом сказываться на результатах при распознавании жестов и других задачах. Эти дефекты не удается удалить с помощью морфологической фильтрации – заполнение больших пустот внутри силуэта руки потребовало бы очень большого структурирующего элемента, что, несомненно, привело бы к разрушению наружного силуэта руки. В настоящей работе эта проблема была решена путем селективного заполнения внутренних контуров на силуэте руки с предварительной обработкой бинарного изображения с помощью морфологических операций.

Для достижения поставленной цели на области сегментированного изображения руки обнаруживаются внутренние контуры, которые заполняются метками руки. Для этого проис-

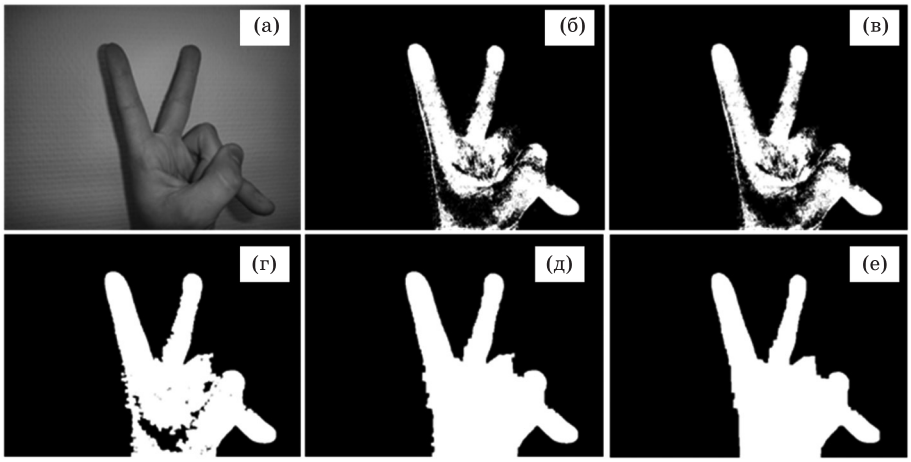

Рис. 4. Исходное изображение – а, б – результат работы цветового фильтра, в – результат удаления шума (фильтр “Б” был выбран автоматически), г – результат операций CLOSE и DILATE, д – результат заполнения контуров, е – результат операции ERODE.

ходит обнаружение внутренних связных компонент на самом большом “белом” связном компоненте на бинарном изображении. Для поиска связных компонент используют двухшаговый алгоритм. Предварительно, перед обнаружением внутренних контуров, последовательно выполняют операции DILATE и CLOSE (рис. 4г), которые позволяют замкнуть слабо разорванные контуры (находящиеся близко к области фона), а также исключить малые пустоты. Поскольку внутренние контуры могут появляться вследствие соединения пальцев руки, то заполняются не все контуры, а только отвечающие одному из следующих условий:

1) площадь найденного контура меньше чем 2,5% от площади руки,

2) компактность (площадь / периметр) найденного контура больше заданного числа.

Использование этих критериев возможно, так как маленькие пустоты и пустоты некомпактной формы, появившиеся в результате соединения пальцев руки, не являются дискриминантными при распознавании жестов и, соответственно, их потеря не приводит к ухудшению результатов распознавания.

После выполнения заполнения внутренних контуров для восстановления исходных размеров силуэта выполняется морфологическая операция ERODE с тем же структурирующим элементом, что использовался при операции DILATE перед этим.

В настоящей работе заполнение внутренних контуров осуществляется два раза. Первый раз – после удаления шума (для того чтобы заполнить внутренние контуры, которые могут быть разрушены последующей морфологической обработкой). Второй раз – после морфологической обработки (для того чтобы заполнить внутренние контуры, вновь образовавшиеся в результате морфологической фильтрации).

Восстановление силуэта руки

Морфологическая обработка изображения ладони с достаточно близко расположенными пальцами может приводить к слиянию изображений пальцев (закрытию промежутков между пальцами). Это характерно для использования структурирующих элементов большого размера при подавлении шумовых включений большой протяженности. В настоящей работе эта проблема решается путем автоматического выбора формы структурирующего элемента.

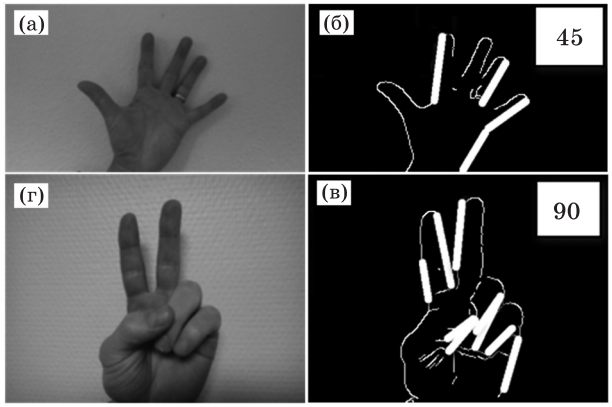

Рис. 5. Исходные изображения – а, б, в, г – результат выделения контуров, широкими полосами обозначены обнаруженные отрезки прямых линий с помощью преобразования Хафа, а в белом квадрате дана автоматически полученная оценка ориентации пальцев руки.

Для этого метод морфологической фильтрации выделенных изображений ладони дополнен процедурой автоматического выбора направления наибольшего удлинения структурирующего элемента путем автоматической оценки направления пальцев ладони и интервалов между ними. Полученная информация используется для ориентирования структурирующего элемента в направлении пальцев. Применение такой адаптации к изображениям всех пальцев руки возможно, так как на руке пальцы либо приблизительно параллельны, либо, если есть сильно повернутые пальцы (например, большой) относительно других пальцев, то они разделены большими интервалами, и ошибка оценивания их ориентации по сравнению с ориентацией других пальцев не приводит к деградации результатов морфологической фильтрации.

Оценка направления пальцев проводится с помощью следующей процедуры:

1. В окрестности ладони полутонового изображения строятся контуры ладони с помощью детектора контуров *Canny* [6].

2. Вычисляется преобразование Хафа (модифицированного для обнаружения отрезков прямых линий) от контурного изображения, построенного согласно п. 1 (рис. 5).

3. Строится гистограмма направления отрезков линий с учетом их длины (отрезки очень малой длины исключаются). Гистограмма имеет 4 ячейки, которые соответствуют углам

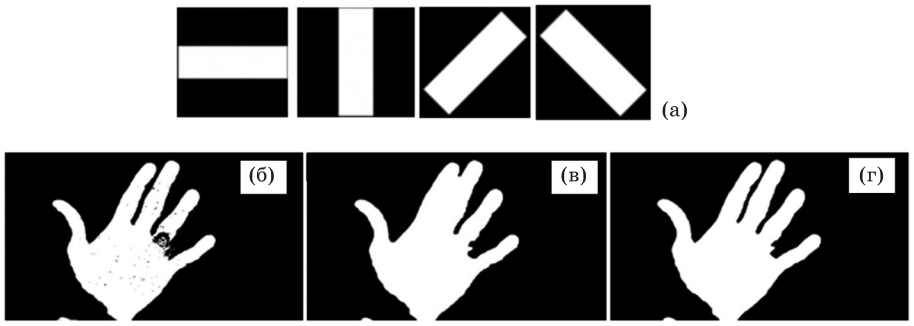

Рис. 6. Ориентированные структурирующие элементы и результаты их применения. Использованные ориентированные структурирующие элементы – а, б – маска, полученная с помощью цветового фильтра, в – результат морфологической фильтрации при использовании квадратного структурирующего элемента, г – результат фильтрации при использовании прямоугольного ориентированного структурирующего элемента.

[0–22,5, 157,5–180], [22,5–67,5], [67,5–112,5], [112,5–157,5]. Ячейка с наибольшим значением определяет направление ладони на изображении (0, 45, 90 и 135 ° соответственно).

4. Формируется структурирующий элемент для морфологической фильтрации – прямоугольник с соотношением длин сторон 1:3, ориентированный длинной стороной вдоль направления, выбранного согласно п. 3 (рис. 6а).

5. Выполняется морфологическая операция CLOSE для бинарной маски, полученной после выполнения процедуры удаления шума согласно алгоритму, описанному выше.

Эксперименты показали, что такой способ оценки ориентации является достаточно устойчивым и позволяет в большинстве случаев выбирать ориентированный структурирующий элемент, который дает результат не хуже и часто лучше чем неориентированный элемент. Эксперименты также показали, что соотношение сторон 3:1 дает наилучший результат при восстановлении силуэтов.

Практические результаты

Для оценки качества работы разработанных алгоритмов были рассмотрены 460 снимков рук разных людей на разных фонах. Используя эту базу данных и цветовой фильтр, принцип работы которого был кратко описан выше, было получено 460 масок изображений. Эти маски были обработаны с помощью морфологической фильтрации согласно схеме, представленной на рис. 2. Разработанная процедура позволила сильно улучшить исходные маски: были убраны шумовые эффекты, соединены “разорванные” сегменты руки – разрывы силуэта, вызванные украшениями рук (кольцами) и ошибками сегментации. На некоторых изображениях можно было заметить отрицательные эффекты. Так, например, на нескольких изображениях ошибочно было заполнено пространство, образованное соединением пальцев, характерным для определенного жеста.

Для более объективной оценки воздействия предложенных алгоритмов на силуэт руки с точки зрения возможности дальнейшего обнаружения, слежения и распознавания жеста на 40 изображениях из базы данных были вручную промаркированы пикселы, принадлежащие изображению руки.

Для оценки качества работы цветового фильтра была посчитана вероятность неправильного принятия решения о принадлежности пиксела изображению руки. Чем ниже эта вероятность, тем выше оценивается качество работы цветового фильтра. В результате морфологической обработки на 40 маркированных изображениях эта вероятность снизилась на 31%. Это косвенно свидетельствует о том, что вероятность успешного обнаружения и слежения по силуэту руки после применения морфологической фильтрации возрастает.

Однако также необходимо оценить, как изменилось количество информации, содержащейся в силуэте руки, которая необходима для распознавания жеста. Можно говорить о том, что наибольшее количество полезной информация содержится на границе силуэта. Таким образом, точность определения границы силуэта является вторым оцениваемым параметром. Для того чтобы получить такую оценку, с помощью изображений, промаркированных

вручную, были получены эталонные контуры силуэтов руки. Контуры силуэтов, полученных с помощью цветового фильтра до и после морфологической обработки, были сравнены с эталонными контурами. Было измерено расстояние от каждого пиксела контура полученных силуэтов до ближайшего пиксела контуров эталонных силуэтов. Результаты приведены на рис. 7.

Можно видеть, что в результате морфологической обработки вероятность корректного обнаружения контуров силуэта значительно увеличивается. Это, в принципе, очевидный результат, поскольку изображения, полученные с помощью цветового фильтра, зашумлены и из-за этого имеют намного большее количество контуров в областях фона и силуэта чем обработанные изображения. Однако по графику, изображенному на рисунке 7б, также видно, что общее количество элементов контуров, совпадающих с контурами эталонных силуэтов, уменьшилось всего лишь на 3%, что говорит о том, что морфологическая фильтрация проведена очень аккуратно и не смещает контуры, которые соответствуют истинным границам силуэта руки.

Для реализации программы была использована библиотека OpenCV. Морфологическая обработка выполнялась на разрешении 640×480, направление руки оценивалось на изображениях, уменьшенных до разрешения 256×192. Время выполнения всей процедуры обработки бинарных изображений зависит от того, насколько сильно раздроблен силуэт, полученный с помощью цветового фильтра. На платформе Intel Pentium IV с 2,8 ГГц одноядерным процессором выполнение процедуры занимает в среднем 23 мс.

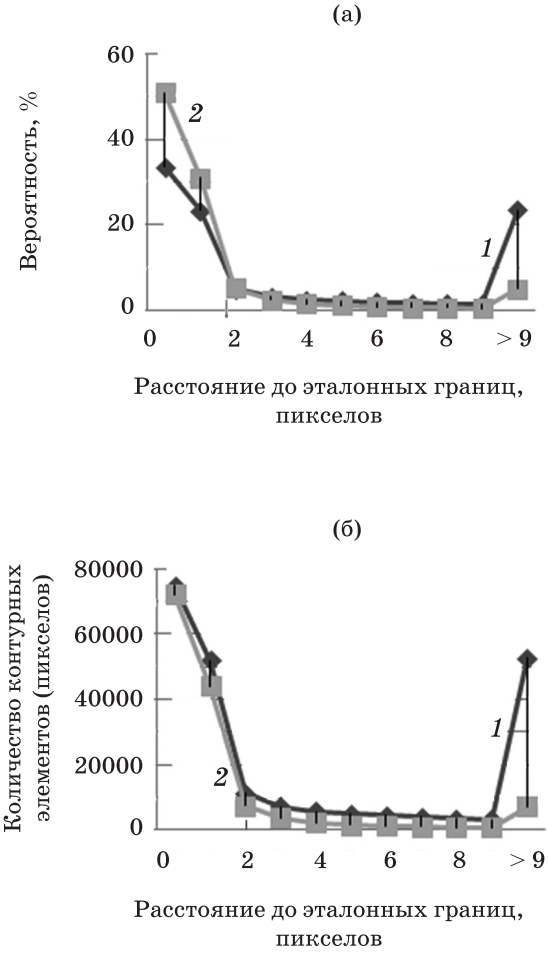

Рис. 7. Распределение вероятности смещения контура сгенерированного силуэта от истинного контура руки (а), б – количество элементов контуров, полученных по сгенерированному силуэту и находящихся на определенном расстоянии от эталонных контуров. *1* – до морфологической обработки, *2* – после морфологической обработки.

Выводы

Рассмотренные алгоритмы обработки изображений силуэтов рук позволяют исправить многочисленные одиночные и групповые ошибки маркирования, полученные в результате применения цветового фильтра для обнаружения пикселов цвета кожи. Были получены косвенные свидетельства того, что использование предлагаемых алгоритмов увеличивает вероятность успешного обнаружения, слежения за рукой, а также распознавания жеста. Отдельно стоит отметить, что применение такой обработки бинарного изображения при незначительной адаптации возможно и при других способах сегментации изображений, например, с помощью глубины или анализа движения. Кроме того, некоторые этапы обработки бинарных изображений применимы не только в задаче распознавания жестов, но и к другим задачам технического зрения, где бинарные маски имеют схожие характеристики.

* * * * *

ЛИТЕРАТУРА

1. *Dente E., Bharath A., Ng J., Vrij A., Mann S., A. Bull A.* Tracking hand and finger movements for behaviour analysis // Pattern Recognition Letters. 2006. V. 27. P. 1797–1808.
2. *Kelly D., McDonald J., Markham C.* A person independent system for recognition of hand postures used in sign language // Pattern Recognition Letters. 2010. V. 31. P. 1359–1368.
3. *Caridakis G., Karpouzis K., Drosopoulos A., Kollias S.* SOMM: Self organizing Markov map for gesture recognition // Pattern Recognition Letters. 2010. V. 31. P. 52–59.
4. *Ben Henia O., Hariti M., Bouakaz S.* A Two-step minimization algorithm for model-based handtracking // 18th Intern. Conf. Computer Graphics, Visualization and Computer Vision (WSCG). University of West Bohemia, Campus-Bory, Plzen, Czech Republic, February 1–4, 2010. P. 189–197.
5. *Oikonomidis I., Kyriazis N., Argyros A.A.* Efficient model-based 3D tracking of hand articulations using Kinect // Proc. 22nd British Machine Vision Conference, BMVC 2011. University of Dundee, UK; Aug. 29–Sep. 1, 2011. P. 101.1–101.11.
6. *Шапиро Л., Стокман Дж.* Компьютерное зрение. М.: Бином. Лаборатория знаний. 2006. 752 с.
7. *Soriano M., Martinkauppi B., Huovinen S., Laaksonen M.* Using the skin locus to cope with changing illumination conditions in color-based face tracking // Proc. IEEE Nordic Signal Proc. Symp. Kolmarden, Sweden. 2000. P. 383–386.
8. *Chiang C., Tai W., Yang M., Huang Y., Huang C.* A novel method for detecting lips, eyes and faces in real time // Real time imaging. 2003. V. 9. № 4. P. 277–287.
9. *Albiol A., Torres L., Delp. E.J.* Optimum color spaces for skin detection // Proc. 2001 Intern. Conf. Image Processing. 2001. V. 1. P. 122–124.