

УДК:681.3.01:621.397.3

Е.В. Малахов, канд. техн. наук, доцент  
А.С. Ерошенко, инженер-программистСЕМАНТИЧЕСКИЙ АНАЛИЗ ИЗОБРАЖЕНИЙ В ЗАДАЧАХ  
СЖАТИЯ ВИДЕОПОСЛЕДОВАТЕЛЬНОСТЕЙ.

У статті розглянуто декілька підходів до стиску та передавання відеопослідовностей. Зокрема: виділення псевдооб'єктів на підставі векторів руху та розшарування статичних зображень. Наведено шляхи подальшого розвитку цих підходів.

Some approaches for compression and videosequence transmission are considered in given article. Such as pseudoobject selection using motion vectors and static image dividing into layers. The further development ways of this methods are shown.

В последнее время в связи с бурным развитием компьютерной техники, Интернета и робототехники значительно возрос интерес к системам понимания изображений. В данной статье речь пойдет о системах понимания изображений предназначенных для анализа изображений, представленных массивом чисел, и составления описания изображенной сцены в «неизобразительной» форме. Характер описания сцены зависит от конкретной задачи. В некоторых случаях желательно чтобы описание было просто указанием на наличие или отсутствие в сцене указанного объекта или появление неожиданного объекта, например такой подход полезен для использования в охранных системах. Или, возможно, нам необходимо отследить движение и направление объекта, это может использоваться в системах наблюдения. В более сложном случае такая система может создавать общее описание сцены, основные элементы сцены и какие объекты в ней находятся. Уже созданы системы, выполняющие относительно простые задачи описания изображений ограниченного класса. В настоящее время исследования направлены на развитие обобщенных программируемых систем, которые могли бы обрабатывать широкий класс изображений, возникающих в разнообразных задачах, в том числе и сжатии видео.

Для систем понимания изображений было предложено множество разнообразных моделей [1, 2]. Исследование этих моделей показывает что в их состав входит одинаковый набор блоков.

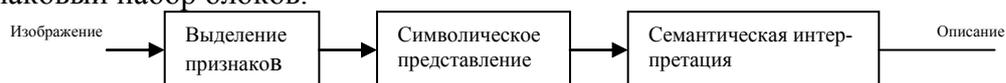


Рис. 1. Обобщенная модель систем понимания изображений.

На рис.1 представлена блок-схема иерархической модели системы понимания изображений. На первом этапе из наблюдаемого изображения **выделяются простейшие признаки**, Признаки могут быть естественными в том смысле, что они устанавливаются визуальным анализом изображения, тогда как другие, искусственные признаки получаются в результате его специальной обработки или измерений. К естественным признакам можно отнести такие признаки как яркость элементов изображения, координаты контурных точек, форма объектов и т.д. Для выделения этих признаков используются различные методы выделения контуров, бинаризации и N-аризации изображений. К искусственным признакам относятся гистограммы распределения яркости, спектры пространственных частот.

Затем этот набор признаков поступает в блок **символического представления**, который формирует символы из признаков. Например, контурные точки группи-

руются в отрезки линий или замкнутые кривые. Элементы с одинаковой яркостью или текстурой объединяются в области, характеризующиеся однородностью.

Множество символов обрабатывается затем блоком **семантической интерпретации**, в результате чего появляется некоторое желаемое описание сцены.

Процесс анализа сцены в большинстве случаев предваряется процессом упрощения – сложный объект (исходное изображение) преобразуется в более простой путем применения нескольких последовательных операций. Простые объекты легче анализировать.

В данной статье анализ сцен рассматривается с точки зрения применения в анализе последовательности видеок кадров с детальным анализом движения всех объектов сцены. Что из себя представляет сжатие видео? Кадры внутри одной сцены обладают явной избыточностью, некоторые области стоят на месте, некоторые движутся, некоторые поворачиваются, но это все одни и те же объекты. Идея состоит в том, чтобы сохранить только опорные кадры сцены, то есть кадры в которых происходят существенные изменения, а все остальные кадры кодируются ссылками на элементы в опорных кадрах. То есть производится поиск элементов из анализируемого кадра в следующем, сравнивая их попиксельно, если найденный объект удовлетворяет некоторой заданной мере сходства, то смещение найдено.

Анализ сцен (движения) можно выполнять на основе движения пикселей, блоков, либо объектов. Так как пиксель является минимальной составляющей изображения, то их использование было бы наиболее приемлемым. Но, к сожалению, информации об одном пикселе недостаточно для нахождения точного смещения именно этого пикселя с достаточно большой долей вероятности. Переход к реальным объектам сцены для решения этой задачи является перспективным, но на данный момент мы не располагаем хорошими и быстрыми алгоритмами выделения объектов. Поэтому при анализе воспользуемся еще одним упрощением: под объектами будем понимать подмножество блоков, обладающих некоторым набором одинаковых или близких по значению свойств. В свою очередь блок – это множество смежных пикселей, причем целое число одинаковых по форме и размеру таких блоков составляет изображение.

В качестве примера для статьи использовались блоки размером 16x16 пикселей. Но, кроме этого, рассматривались блоки размером 8x8 и 4x4. Наиболее оптимальным оказалось использование блоков 8x8, так как при использовании блоков 4x4 в блоке недостаточно информации для адекватной оценки степени сходства блоков.

Данная работа основана на изучении смещений блоков в полярных координатах, то есть мы при исследовании будем оперировать двумя координатами: фазой и длиной вектора смещения блока.

Цель такого подхода следующая. Предположим, что блок движется в одном направлении несколько кадров подряд, координаты и, соответственно, расстояния его смещений будут всегда разные, но фаза останется такой же. Это дает ряд преимуществ. В частности, появляется возможность выделить области движущиеся в одном направлении, и, как следствие, использовать вектор движения как один из критериев включения блоков в объект, то есть выделения движущегося объекта и его направления. Таким образом, можно рассматривать фильм как совокупность объектов и при воспроизведении передавать полную информацию о каждом объекте только один раз, а затем только координаты положения объектов или вектора их смещений.

Кроме того, этот подход может быть полезен в Motion Estimation (ME). При записи векторов в полярном виде можно использовать факт выделения больших областей с одинаковым вектором смещения, что приводит к уменьшению количества информа-

ции, необходимой для описания смещения изображения и, соответственно, к лучшему сжатию. Однако, может возникнуть ситуация, когда для хранения смещения потребуются хранить числа с плавающей точкой, что при большом количестве и малых размерах таких областей может привести не к сжатию, а наоборот, к увеличению выходных данных.

Использование данного подхода в МЕ заключается в следующем. Изображение разбивается на блоки заданного размера. Затем вычисляются оптимальные смещения блоков. Критерием оптимальности является величина ошибки. Если ошибка, например MSE, достаточно мала, то считается, что блок действительно сместился на полученный вектор. Если ошибка большая, то блок нужно перекодировать или разбивать на более мелкие блоки. Затем для сместившихся блоков вычисляем фазу вектора смещения (например, в градусах от  $0^\circ$  до  $359^\circ$ ). (Рис.1) и вносим полученные результаты в отдельную матрицу. Если ошибка для этого блока больше заданного порога, то в матрицу вносим значение 360. Одновременно вычисляем матрицу длин векторов, в которой хранится абсолютное расстояние между блоками в пикселях.

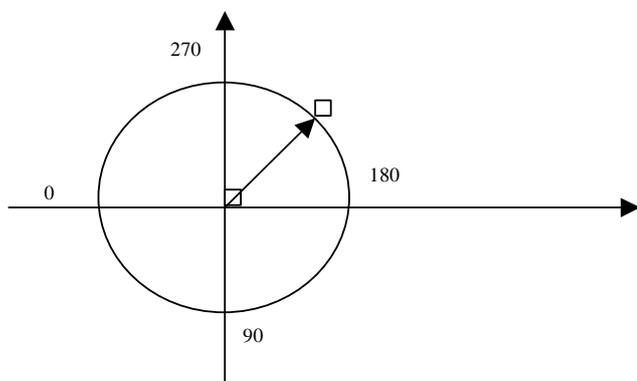


Рис. 2. Смещение блока в полярных координатах.

На Рис.2 приведен пример двух соседних кадров из тестовой видеопоследовательности «Робот».



Рис. 3. Кадры из видеопоследовательности «Робот» (Сцена 1).

В кадрах наблюдается небольшое смещение сцены вправо.

При анализе углов смещений блоков и смещений блоков получается следующая матрица длин и фаз векторов (Рис. 4).

0	0	180	270	180	360	180	360	181	0	359	360	180	360	180	359	360	0	180	359	180	180	1
0	180	180	180	180	180	180	180	0	180	180	180	180	0	180	0	180	180	0	0	0	0	2
360	360	180	180	180	180	180	270	180	180	0	180	180	360	360	360	0	180	0	359	180	0	3
0	180	180	180	180	180	180	180	180	180	180	360	180	360	360	360	180	180	180	360	360	360	4
360	360	180	180	180	180	180	180	180	180	180	180	180	180	180	360	360	360	360	360	360	360	5
360	180	180	360	180	180	180	180	180	180	180	180	180	180	180	180	180	360	180	180	360	360	6
0	180	180	180	180	180	360	360	180	180	180	180	180	180	180	180	180	180	180	180	180	360	7
180	180	180	180	180	360	360	360	180	180	180	180	180	180	180	180	180	180	180	180	180	360	8
360	360	360	360	180	180	360	360	180	180	180	180	180	180	180	180	180	180	180	180	180	360	9
360	360	360	180	360	180	360	360	360	180	180	180	180	180	180	180	180	180	180	180	180	360	10
360	360	360	360	360	180	180	180	180	360	180	180	180	180	180	180	180	180	180	180	180	360	11
360	360	360	360	360	180	180	180	180	360	360	180	180	180	180	180	180	180	180	180	180	360	12
360	360	360	360	180	180	180	180	180	180	360	359	180	180	180	180	180	180	180	180	180	360	13
359	180	180	180	180	180	180	180	180	180	180	0	180	180	180	180	180	180	180	180	180	0	14
180	0	179	180	180	0	180	180	0	180	0	0	180	178	180	180	180	0	180	180	0	180	15
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	

Рис. 4. Матрица смещений векторов для видеопоследовательности «Робот» (Сцена 1).

Из рисунка видно, что кадр можно разделить на небольшое число областей, каждая из которых движется в одном направлении с одним смещением. Одна из областей стоит на месте, другие две движутся со смещением 3 и 4 пикселя в указанном в таблице направлении. Если вектор смещения блока считать одним из его свойств, то, в соответствии с приведенным ранее упрощением понятия объекта, все выделенные области можно считать объектами в этом смысле.

Следующим шагом кадр можно разделить на несколько слоев, один слой – один движущийся объект. При воспроизведении будет достаточно передавать только координаты движения слоев с их последующим наложением.

Приведенный на Рис.2. и Рис. 3 пример характерен еще с одной точки зрения. В этой сцене есть области (объекты), имеющие значительный размер по отношению ко всему изображению. Этот факт свидетельствует не столько о движении объектов в кадре, сколько о движении камеры относительно объектов или фона сцены. (Рис.5).

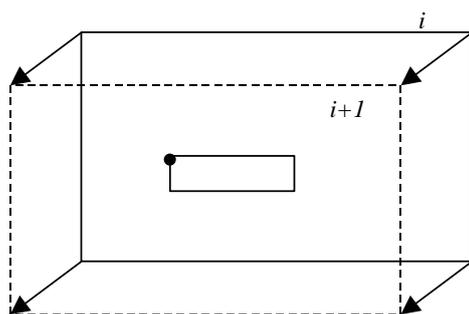


Рис. 5. Движение камеры относительно объектов или фона сцены.

Если при этих условиях выполнить сдвиг изображения на вектор области, покрывающей наибольшую его часть (Рис.6), то вектора остальных объектов (слоев) можно пересчитать относительно нового начала координат.

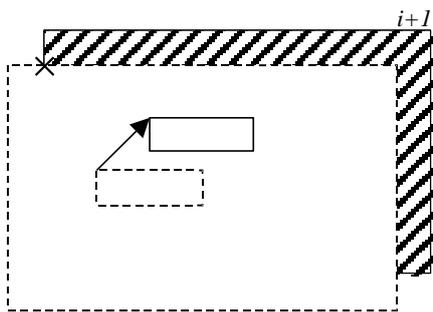


Рис. 6. Сдвиг изображения на вектор наибольшей области.

Очевидно, что часть изображения, не покрытую наибольшей областью, необходимо будет закодировать заново.

Рассмотрим следующую сцену (Рис.7).



Рис. 7. Кадры из видеопоследовательности «Робот» (Сцена 2).

Смещения блоков, полученные в этих кадрах, показаны на Рис.8.

359	180	360	359	180	360	181	180	0	181	180	0	359	180	180	360	359	180	359	358	180	270	1	
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	178	0	0	0	0	0	0	0	3
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5
0	0	0	0	0	179	180	0	180	180	180	180	0	0	0	0	0	0	0	0	0	0	0	6
0	0	0	0	0	180	180	0	0	0	180	0	0	0	0	0	0	0	0	0	0	0	0	7
0	0	0	0	0	0	360	360	360	0	180	180	180	0	0	0	0	0	0	0	0	0	0	8
0	0	0	0	0	360	0	360	360	0	180	180	180	180	0	0	0	0	0	0	0	0	0	9
0	0	0	0	0	0	0	360	360	0	0	0	180	180	0	0	0	0	0	0	0	0	0	10
0	0	0	0	0	0	0	0	360	0	0	0	0	180	0	0	0	0	0	0	0	0	0	11
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	12
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	13
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	14
0	0	179	180	180	180	1	0	179	180	0	180	1	180	0	1	0	180	178	180	180	180	180	15
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22		

Рис. 8 Смещения блоков в кадрах видеопоследовательности «Робот» (Сцена 2).

В этой сцене четко видна большая статическая область, которая не перемещается. Женщина с зонтиком вся сместилась на 1 пиксель, теперь рассмотрим что происходит с роботом. Нижняя половина робота вся перекодируется, по видимому в этих блоках произошли большие изменения. В верхней части, там где голова робота вектора смещений всех блоков обладают одинаковой фазой, но разной длиной. Анализируя эту ситуацию на изображении видим поворот реального объекта, но так как мы имеем дело с двумерным объектом, назвать это поворотом не корректно. Это можно назвать деформацией объекта. Вероятнее всего эта деформация подчиняется некоторому закону, и следовательно, этот закон можно найти и использовать для уменьшения количества описательной информации.

Интересные результаты получены на кадрах, где происходит увеличение или уменьшение всей сцены или некоторых объектов. В этом случае блоки, находящиеся в фокусе приближения или удаления, стоят на месте, а направление остальных стремится к фокусу или от фокуса.

Для разделения изображения на слои и их дальнейшего использования в случае увеличения или уменьшения объекта на изображении необходимо выполнить предварительный анализ сцены или её части с целью определения кадра с минимальным размером такого объекта (Рис. 9). На этом кадре осуществить расслоение с целью получения слоя с максимальной видимой областью фона (заднего плана). Именно этот слой будет передан единожды.

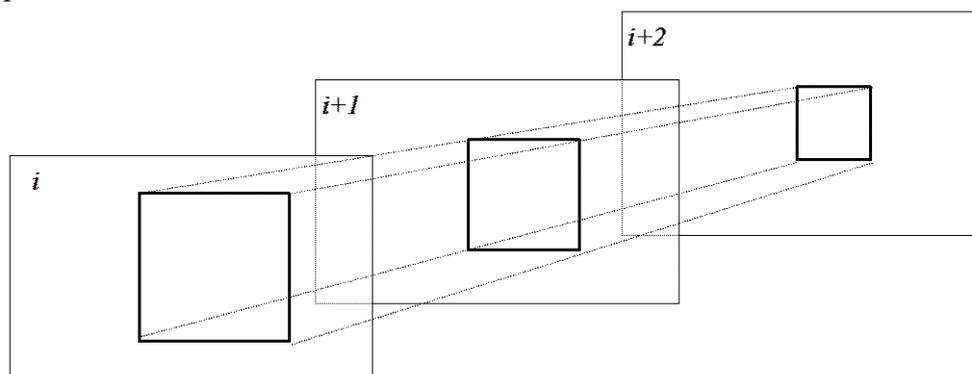


Рис. 9 Масштабирование (уменьшение) объекта сцены.

Слой с объектом максимального размера также передается один раз с указанием размера (коэффициента масштабирования и координат местоположения) в  $i$ -м кадре. Затем передается аналогичная информация для  $(i+1)$ -го кадра. Недостатком является то, что данный подход неприменим для кодирования в реальном масштабе времени.

И наконец, еще один момент, характерный для многих фильмов. Это наличие незначительной флуктуации объектов сцены или изображения в целом, вызванной различными причинами: дрожанием камеры, движением воздуха при съемке, волнением водной поверхности и т. п. Если эта ситуация не является художественным замыслом автора видеофрагмента, то ее можно отнести к категории шума и попытаться каким-либо образом устранить.

Одним из вариантов решения данной задачи может быть использование описанного подхода оценки движения в полярных координатах.

В данном случае речь идет о фильтрации путем огрубления векторов движения. Например, можно использовать не точное значение фазы вектора (Рис. 2), а усредненное значение в пределах  $\alpha=3 - 5^\circ$  (Рис. 10).

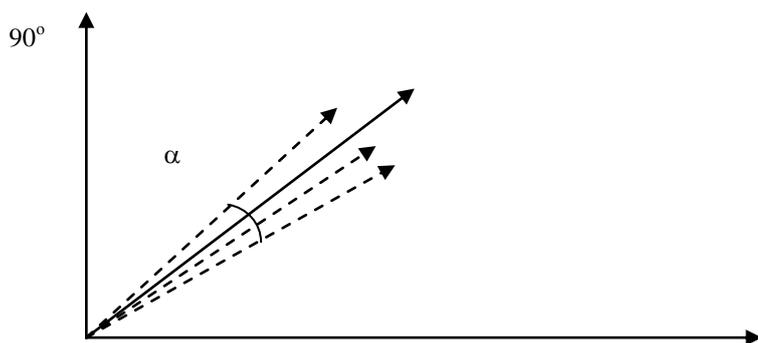


Рис. 10. Усреднение значения вектора в заданных пределах.

Такой подход имеет много преимуществ для анализа фильма, в частности это позволит сократить количество передаваемой информации. Например, можно ограничиться краткой информацией о движении блоков и сегментов (подмножества блоков). С этой целью проводится предварительная классификация сцен:

1. Сцена с плавным движением в одну сторону  
В этом варианте можно провести декомпозицию кадра на передний и задний план (возможно, на большее количество планов), а также определить размер планов и о том, на сколько пикселей смещается каждый план. Предполагается, что движение блоков в сцене подчиняется некоторому закону, в этом случае движение блоков можно описать одним или несколькими числами.
2. Сцена с плавным приближением или удалением всей сцены  
Проводится анализ сцены на приближение или удаление, центр фокуса, его размер, далее необходимо изучить закон движения блоков, возможно, его также можно задать одним или несколькими параметрами.
3. Более сложные сцены, включающие некоторую комбинацию предыдущих, в том числе сцены включающие повороты/деформации объектов.

В дальнейшем, для реализации такого подхода необходимо привлечь дополнительно аппарат выделения объектов, так как оперирование блоками не является естественным. Одним из вариантов может быть бинаризация изображения, N-аризация, а также выделение объекта по его векторам смещения, а затем для уточнения границ производить анализ блоков на краях объекта.

### Список литературы

1. Р. Дуда, П. Харт «Распознавание образов и анализ сцен».
2. У. Претт «Цифровая обработка изображений».