

Введение

Разработанный метод сочетает в себе две стратегии сравнения изображений внутри серии: на основе гистограмм и на основе поиска общих локальных особенностей между изображениями.

Метод, основывающийся на гистограммах, дает основную часть ответов о том, какие изображения составляют панораму (рейтинг более 96% на первой базе изображений). После подбора параметров этого метода (выбор цветового пространства, дистанции и т.д.) для получения максимального качества были отброшены многие другие подходы, т.к. они давали те же ответы, что и гистограммная стратегия. В связи с этим, мы потратили много времени на поиск другого алгоритма, который давал бы правильные соответствия между изображениями, в случаях, которые не решаются подходом с использованием гистограмм.

Наш выбор остановился на алгоритме нахождения соответствий между локальными особенностями. Опять же нам пришлось очень долго подбирать параметры этого алгоритма, так как мы столкнулись с большим количеством проблем на текущей базе изображений:

- небольшое разрешение изображений, что влечет за собой низкую повторяемость (repeatability) детектора особенностей;
- наличие больших не текстурных областей (асфальт, небо, ...) на изображении;
- низкая текстурность объектов, присутствующих на сцене.

Для решения проблемы повторяемости нам пришлось подбирать параметры детектора таким образом, чтобы он выдавал большое количество особенностей на каждом изображении (~10к). Это повлекло очень большие вычислительные затраты при нахождении соответствий между дескрипторами, вследствие чего мы стали использовать FLANN подход. Это привнесло случайность в наши результаты, однако позволило уложиться в сутки, которые отводились для получения результата на финальной базе. Из-за низкой текстурности объектов многие соответствия, которые мы находили, были неправильными. Для нахождения правильного подмножества соответствий был использован алгоритм поиска гомографии между изображениями с использованием схемы RANSAC. Так как доля правильных соответствий была очень низкой, нам пришлось реализовать различные фильтры, позволяющие отбросить неправильные соответствия, чтобы использовать достаточно небольшое количество итераций схемы RANSAC.

Затем нам оставалось только слить результаты этих двух подходов с весами пропорциональными нашей уверенности в них.

Стратегия на основе сравнения гистограмм

Идея использовать гистограммный подход возникла из гипотезы, что изображения из одной панорамы имеют близкие распределения цветов, т.к. отображают одну сцену:

- с определенными условиями освещенности;
- с определенными временем года (цвет листвы);
- с определенными погодными условиями (цвет неба).

При выполнении этих условий можно предполагать схожесть цветовых характеристик и яркости пикселей на изображениях из одной панорамы. В связи с этим можно ввести дистанцию между парой изображений как дистанцию между гистограммами соответствующими этим изображениям. В качестве цветового пространства в итоге было экспериментально выбрано пространство YCrCb.

Мы рассматриваем 4D гистограммы по следующим размерностям:

- интенсивность Y (12 бинов гистограммы)
- цветовые характеристики Cr и Cb (256 бинов)
- ордината пикселей на изображении y (4 бина)

Количество бинов было подобрано экспериментально. Последняя размерность фактически влечет учет пространственного расположения цветов: дорога должна сравниваться с дорогой, небо с небом, трава с травой. Подобный подход позволяет более точно сравнивать подобные типы областей. Необходимо заметить, что разбиение изображения по горизонтальным полосам было введено в связи с тем, что изображения внутри одной панорамы преимущественно сдвинуты относительно друг друга только по горизонтали.

Небольшое количество бинов по интенсивности учитывает тени и некоторые вариации освещенности на картинках.

В качестве дистанции между гистограммами была экспериментально выбрана $CV_COMP_INTERSECT$ (Histogram Intersection Distance).

После вычисления гистограммы каждого изображения в серии, мы строим симметричную матрицу размерности 5 на 5, содержащую дистанции между гистограммами. Элементы на диагонали установим нулевыми, чтобы исключить из рассмотрения пары изображений (I, I) . Эта матрица впоследствии будет использоваться для нахождения панорамы внутри серии изображений.

Стратегия Cross-Check

Эта стратегия основана на поиске и сравнении локальных особенностей изображений (достаточно уникальных участков изображений). Примером таких локальных особенностей могут быть угловые точки, находящиеся на изображении. Применение подобной стратегии для решения поставленной задачи требует следующих предположений:

- 1) Изображения, составляющие панораму, должны иметь перекрытия, то есть для каждого изображения I должно существовать изображение J , содержащее часть объектов, присутствующих на изображении I .
- 2) Для перекрывающихся изображений должно существовать достаточное количество общих локальных особенностей.

Для предложенной базы изображений эти предположения выполняются для большей части входящих в нее панорам. Изображения в тестовой базе достаточно текстурные, поэтому общая область перекрывающихся изображений содержит достаточно большое количество особенностей.

Вход: 5 изображений серии.

Выход: Матрица смежности A размером 5×5 . $A(I, J)$ - некоторая мера уверенности, что изображения I и J принадлежат к одной панораме. В качестве такой меры уверенности в данной стратегии служит количество локальных особенностей изображения I (J), имеющих соответствия на изображении J (I).

Описание алгоритма:

1. Для каждого изображения из серии находим локальные особенности. Для этого запускаем детектор (алгоритм нахождения локальных особенностей) $FAST^1$ на 4 уровнях Гауссовой пирамиды, вычисленной для изображения. Объединяем локальные особенности, найденные на всех уровнях пирамиды изображения.

Мы подобрали такие параметры $FAST$ детектора, чтобы он давал очень много (~ 10000) пусть не совсем уникальных локальных особенностей для изображения (300×300). Это позволило в

¹ <http://mi.eng.cam.ac.uk/~er258/work/fast.html>

случае низкого разрешения повысить повторяемость (repeatability) детектора на паре изображений, имеющих пересечение.

2. Для локальных особенностей каждого изображения вычисляем дескрипторы (описатели). Мы использовали SURF² дескриптор. SURF дескриптор представляет в нашем случае вектор из 64 float-значений.
3. Для каждой пары изображений I ($I=1,...,5$) и J ($J=1,...,5, J \neq I$) находим соответствия между их локальными особенностями. Для этого мы для SURF-дескриптора каждой локальной особенности изображения I находим ближайший (по L1 дистанции) SURF-дескриптор среди локальных особенностей изображения J . Чтобы ускорить нахождение ближайших дескрипторов мы использовали алгоритм, основанный на построении ускоряющих структур на множестве дескрипторов изображения (из библиотеки flann³, являющейся сейчас модулем OpenCV). Заметим, что ускоряющие структуры во flann строятся случайным образом, поэтому результаты работы алгоритма могут отличаться от запуска к запуску.
4. Соответствия для каждой пары изображений I ($I=1,...,5$) и J ($J=I+1,...,5$) фильтруем по критерию надежности Cross-Check: соответствие между дескриптором d_i с изображения I и дескриптором d_j с изображения J считается надежным, если для дескриптора d_i ближайшим (в пространстве дескрипторов) дескриптором на изображении J будет являться дескриптор d_j , а для дескриптора d_j ближайшим на изображении I будет являться дескриптор d_i . В результате этого шага для изображений I и J знаем множество взаимно однозначных соответствий между локальными особенностями.
5. Фильтруем оставшиеся соответствия по y координате: если y -координаты точек, связанных соответствием, отличаются больше чем на 100 пикселей, то такое соответствие считаем ложным. Этот фильтр обоснован тем, что в конкурсных изображениях встречались только горизонтальные панорамы.
6. Для каждой пары изображений I ($I=1,...,5$) и J ($J=I+1,...,5$) по множеству взаимно однозначных соответствий между локальными особенностями вычисляем матрицу гомографии⁴ по схеме RANSAC⁵. Согласно теории, преобразование гомографии существует для точек изображений (при произвольном изменении положения камеры), если их прообразы в 3D мире лежат на одной плоскости. Для неплоских объектов мы также можем искать преобразование гомографии, если отклонение точек объекта от «плоской» формы много меньше, чем расстояние от камеры до объекта. Поиск гомографии для конкурсных изображений панорам обоснован следующими фактами: 1) для объектов городских сцен характерно наличие плоскостей; 2) объекты часто достаточно удалены, чтобы считать их плоскими.

В цикле поиска гомографии по схеме RANSAC для очередной гипотезы о гомографии применялись фильтры:

- 1) Если минимальное сингулярное число для матрицы гомографии меньше некоторого порога, то такую гомографию мы не рассматриваем, поскольку она близка к вырожденной и сводит все (многие) точки одного изображения в одну точку другого изображения.
- 2) К соответствиям, удовлетворяющим текущей гипотезе о гомографии (inliers, инляерам), применяются фильтры:
 - а. Orientation consistence check. Локальные особенности изображения имеют несколько свойств: положение, размер, ориентация и т. д. Данный фильтр находит

² <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.85.2512&rep=rep1&type=pdf>

³ <http://www.cs.ubc.ca/~mariusm/index.php/FLANN/FLANN>

⁴ <http://en.wikipedia.org/wiki/Homography>

⁵ <http://ru.wikipedia.org/wiki/RANSAC>

для инлаеров преобладающее изменение ориентации и фильтрует соответствия, соединяющие локальные особенности, изменение ориентации которых отклоняется от преобладающего изменения больше, чем на заданный порог.

- b. Фильтр «гантель». «Гантелей» мы назвали ситуацию, когда на одном изображении есть небольшая область (радиуса 11), в которой есть больше 2 точек (центров локальных особенностей), имеющих соответствия с точками, находящимися в такой же маленькой области другого изображения и при этом в окрестности большего радиуса (=33) других точек почти нет (по порогу). Обнаружив «гантель», мы заменяем множество соответствий, ее составляющих, на одно соответствие.
- c. Геометрический фильтр. Выполняется по углу соответствий. Угол соответствия, соединяющего точку (x_i, y_i) (in $[1, 300] \times [1, 300]$) изображения I и точку (x_j, y_j) (in $[1, 300] \times [1, 300]$) изображения J, определяется как угол наклона вектора из точки (x_i, y_i) в точку $(x_j + 300, y_j)$. По всем соответствиям между изображениями I и J находится доминирующее направление и среднестатистическое отклонение от него. Фильтруются соответствия, которые отклоняются от доминирующего, больше чем на среднестатистическое отклонение.

Мы считаем, что мы нашли «правильную» гомографию между изображениями, если ей удовлетворяет больше 100 инлаеров (после фильтров а.) - в.)). Если такой гомографии не нашлось за заданное число итераций RANSAC, то выдаем ту гомографию из всех рассмотренных, которой удовлетворяет наибольшее число соответствий.

Заметим, что схема RANSAC выбирает 4 точки для вычисления гомографии случайным образом, поэтому результаты работы программы могут отличаться от запуска к запуску.

- 7. Заполняем матрицу смежности A (5x5). $A(I, J)$ — число инлаеров, найденных между изображениями I и J.

Методы смешивания стратегий

Для вычисления финальной гипотезы о подмножестве серии, содержащем панораму, мы сведем нашу задачу к задаче на графе. Будем рассматривать 5 изображений серии как вершины графа, соединенные неориентированными ребрами с весами, вычисляемыми на основе количества правильных соответствий, найденных с помощью алгоритма CrossCheck, и на основе дистанций между гистограммами.

Алгоритм вычисления весов ребер следующий:

1) Гистограммы:

- После вычисления дистанций между гистограммами, мы получаем матрицу M размерности 5*5, содержащую веса ребер. Большие значения дистанции соответствуют большей вероятности ребра между вершинами.
- Чтобы использовать алгоритм Краскала, предварительно изменим матрицу следующим образом:
 $M := 300 \times 300 - M$.

Теперь меньшее значение дистанции соответствует большей вероятности ребра.

- После этого нормализуем полученные дистанции таким образом, чтобы они попадали в интервал значений $[0,1]$. Для этого введем три порога уверенности: absoluteConfidence, randomConfidence, zeroConfidence (их значения подбираются экспериментально). Все ребра, вес которых меньше absoluteConfidence, мы считаем абсолютно правильными; ребра меньшие по весу чем randomConfidence могут оказаться как правильными, так и неправильными. Ребра, чей вес больше zeroConfidence, мы считаем неправильными. Далее преобразуем каждое значение в матрице с помощью кусочно-линейного преобразования на основе этих параметров:

$$newDistance = \begin{cases} 0, & distance < absoluteConfidence \\ 0.5 * \frac{distance - absoluteConfidence}{randomConfidence - absoluteConfidence}, & absoluteConfidence < distance < randomConfidence \\ 0.5 + 0.5 * \frac{distance - randomConfidence}{zeroConfidence - randomConfidence}, & randomConfidence < distance < zeroConfidence \end{cases}$$

После всех этих преобразований мы получаем матрицу, соответствующую ребрам графа, причем веса ребер могут изменяться от 0 до 1.

2) Cross-Check

- Результатом работы этой стратегии является матрица A, содержащая количество найденных соответствий между соответствующими изображениями. Чем больше элемент матрицы, тем больше найдено соответствий между изображениями, и тем более вероятно это ребро в графе.
- Аналогично стратегии с использованием гистограмм модифицируем матрицу $A := 100000 - A$.
- Аналогично введем кусочно-линейное преобразование, переводящее каждое значение матрицы в интервал $[0,1]$. Коэффициенты преобразования были подобраны экспериментально.

Объединим результаты этих стратегий, сложив полученные матрицы с экспериментально подобранным коэффициентом $F = \min(a * M, A)$. Коэффициент a меньше единицы, тем самым мы придаем больший приоритет результатам работы алгоритма, использующего гистограммы.

Далее используем алгоритм Краскала для нахождения максимального остовного дерева в этом графе, считая, что веса ребер заданы матрицей F:

- Вначале текущее множество рёбер устанавливается пустым.
- Из всех рёбер, добавление которых к уже имеющемуся множеству не вызовет появление в нём цикла, выбирается ребро минимального веса.
- Если вес этого ребра меньше заданного экспериментально выбранного порога $Thresh$, это ребро добавляется к текущему множеству. Иначе, если количество вершин в максимальной связной компоненте графа, содержащего текущее множество ребер, меньше трех, то все равно добавляем его к множеству. Если же размер не меньше трех, то считаем, что все изображения, соответствующие вершинам максимальной связной компоненты полученного графа, составляют единую панораму.

Для подбора параметров на обучающей выборке мы использовали следующий подход. Рассмотрим тренировочную базу из 1000 серий. Для каждой серии построим матрицу весов для стратегии с использованием гистограмм (весом будет дистанция между гистограммами) и стратегии с использованием локальных особенностей (весом будет количество правильных соответствий).

Для каждой стратегии:

- Переберем все веса ребер, встречающиеся в матрицах, соответствующих этой стратегии;
- Для каждого веса посчитаем рейтинг на тренировочной базе, принимая решение о составе панорамы с помощью алгоритма Краскала, основываясь только на матрице весов этой стратегии, и считая, что $Thresh$ равен выбранному весу.
- $randomConfidence$ будем считать равным весу, отвечающему лучшему значению рейтинга.
- Также по аналогичной схеме вычислим порог $absoluteConfidence$ так, чтобы все ребра, меньшие его по весу, были бы правильными, то есть связывали бы изображения внутри одной панорамы.
- Порог $zeroConfidence$ выбирается исходя из аналитических соображений.

После чего нормализуем эти матрицы с использованием найденных параметров кусочно-линейных преобразований. Далее экспериментально был подобран коэффициент α , использующийся для объединения стратегий. Для этого мы перебирали различные значения коэффициента в некотором отрезке, стараясь максимизировать рейтинг, каждый раз находя оптимальный порог алгоритма Краскала для объединенной матрицы весов по аналогичной схеме.