

УДК 519.92

Исследование методов и разработка алгоритмов обработки видеoinформации в задачах локализации положения БЛА на основе распознавания изображений при помехах и искажениях

И.С. Пуртов, Д.П. Синча

Аннотация

В статье описаны основные этапы работы алгоритма локализации местоположения беспилотного летательного аппарата(БЛА), основанного на анализе ключевых признаков изображения и их дальнейшей обработки. Представлены результаты работы алгоритма поиска изображения в графической базе данных, представляющей собой набор тестовых снимков, полученных с бота БЛА.

Ключевые слова

распознавание образов, ключевые признаки изображения, SIFT, SOM, кластеризация, локализация местоположения БЛА.

Введение

Одной из актуальных задач современной навигации является обеспечение автономной высокоточной навигации беспилотных летательных аппаратов (ЛА) в сложных погодных и сезонных условиях над различными ландшафтами. При этом высокоточное определение координат своего местоположения должно осуществляться в реальном масштабе времени.

При регистрации фотографии с борта БЛА на изображения накладываются различные помехи, что усложняет дальнейших процесс обработки данных. Любое изображение в отдельности, несмотря на кажущуюся простую структуру – двумерную матрицу чисел, содержит очень сложную и комплексную – интегральную информацию о наблюдаемой сцене. Извлечь из изображений какую-нибудь структурированную и осмысленную информацию о наблюдаемой сцене невероятно сложно.

Изображение, полученное с движущейся камеры, особенно если на нем запечатлена динамическая сцена, анализировать еще сложнее. Поэтому просто необходимы технологии, которые позволили бы извлекать из изображений, полученных с борта БЛА, некоторую, пускай

очень неполную, но осмысленную и достаточно просто структурированную информацию об объектах сцены и ее динамике.

Постановка задачи

Задача локализации местоположения БЛА на основе распознавания фотографий подстилающей поверхности может быть сформулирована следующим образом: в качестве начальных данных задается тестовое изображение, полученное с борта БЛА, при этом существует графическая база данных (фотографий местности возможного расположения БЛА), каждый снимок которой имеет четкую координатную привязку.

Подающееся на вход системы изображение может быть искажено одним или несколькими следующими факторами:

- групповые помехи произвольной формы и размера,
- поворот или сдвиг,
- нелинейное изменение яркости,
- изменение разрешения или изменение масштаба,
- кадрирование/ декадрирование,
- зеркальное отображение.

Необходимо найти соответствие между тестовым снимком местности расположения БЛА и снимками, хранящимися в базе данных и определить текущие координаты БЛА на основе полученной информации.

Вследствие наличия искажающих факторов необходимо разрабатывать алгоритм выделения таких признаков изображения, каждый из которых будет малочувствителен хотя бы к одному из рассматриваемых факторов. По этой же причине разрабатываемые признаки изображения должны, по крайней мере, слабо коррелировать, а лучше, если они будут иметь слабую статистическую взаимосвязь.

Необходимо учитывать, что признаковое распознавание изображений может оказаться весьма ресурсоемким из-за необходимости вычисления большого количества признаков и их последующего сравнения. Однако при наличии перспективных искажений или сложных типов изменчивости объектов на изображениях альтернативу признаковым методам найти затруднительно.

Для отображения изображения в дескриптор мы используем модель представления изображения, называемую сумка визуальных слов [1]. Для того чтобы представить изображение множеством (сумкой) визуальных слов необходимо обнаружить на изображении точки интереса (признаки изображения), описать эти точки, а далее заменить фактические описания визуальными

словами, используя словарь визуальных слов. Последний получают посредством кластеризации множества описаний точек интереса всех входных изображений.

Алгоритм SIFT

В качестве алгоритма поиска характерных признаков изображения был выбран алгоритм SIFT, который включает в себя, и алгоритм обнаружения, и алгоритм описания. В качестве точек интереса, обнаруживаемых SIFT-детектором выступают так называемые «пятна» (blobs). Говоря нестрого, пятно – эта область изображения, которая ярче/темнее, чем её окружение (рисунок 1). SIFT-описание пятна (SIFT-дескриптор) обладает привлекательными свойствами, определившими его использование нами. Во-первых, он инвариантен (достаточно устойчив) к повороту, переносу и изменению масштаба изображения, т.е. почти к полной группе аффинных преобразований. По мнению некоторых авторов [2] никаких существенных улучшений SIFT-дескриптора с точки зрения повышения устойчивости к указанным преобразованиям никогда не будет. Во-вторых, он устойчив к изменению освещения сцены. В-третьих, он, устойчив к смене ракурса – вплоть до 60° , если исходное изображение фронтальное (для БЛА – снято в надир). В-четвертых, SIFT-дескриптор характеризуется высокой информационной ёмкостью. Также, по сравнению с другими существующими дескрипторами, он более устойчив к размытию изображения (которое может возникнуть, например, вследствие неверной фокусировки камеры БЛА).

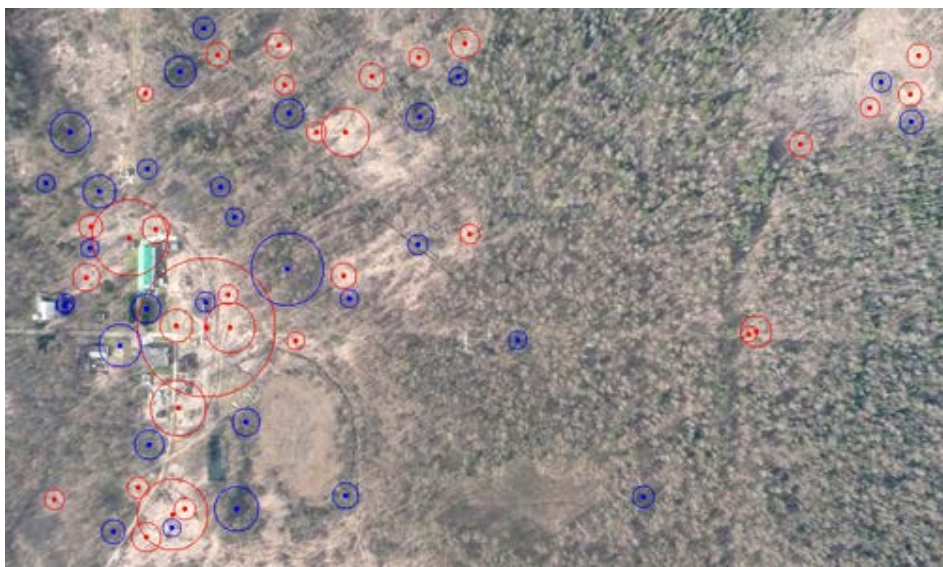


Рисунок 1 – Ключевые признаки изображения и соответствующие им «пятна», выделенные алгоритмом SIFT. Красный цвет соответствует светлым областям, синий – тёмным.

Результаты работы SIFT-детектора, как и многих других существующих, неудовлетворительны для малоконтрастных изображений. Поэтому каждое изображение,

поступающее на вход разработанной системы, проходит этап предварительной обработки, на котором повышается его контрастность.

Кроме того после повышения контрастности для каждого изображения формируется несколько его копий, подвергнутых фиксированному набору различных аффинных преобразований (используется алгоритм ASIFT). Таким образом мы получаем как бы несколько возможных фотографий данного «места» с разных точек обзора. Это существенно расширяет возможности стандартного SIFT-алгоритма.

Алгоритм ASIFT

Если физический объект имеет гладкую или кусочно-гладкую границу, его изображения, полученные с камер, находящихся на разных позициях подвержены гладкой деформации. Эти деформации локально хорошо аппроксимируются аффинными преобразованиями плоскости изображения.

В результате задача распознавания объектов часто приводит к вычислению аффинно-инвариантных локальных признаков изображения. Такие инвариантные функции могут быть получены методами нормализации, но на данный момент таких полноценных методов не существует. Моделируя на обоих изображениях различные масштабы и нормализуя смещение и поворот, SIFT представляет собой метод, полностью инвариантный к четырем из шести параметров аффинного преобразования.

Affine-SIFT (ASIFT), имитирует набор различных видов на объекты с исходных изображений, которые можно получить, варьируя два параметра ориентации оси камеры, а именно углы широты и долготы. Эта процедура не входит в стандартный метод SIFT. Затем он применяет метод SIFT ко всем изображениям, которые генерируются таким образом. В итоге ASIFT охватывает фактически все шесть параметров аффинного преобразования. Математически доказано что этот метод полностью аффинно-инвариантен. Основная анаморфоза (деформация) от изображения к другому, вызванная применением аффинных преобразований, может быть измеренна по *переходу наклона*, новому геометрическому понятию, описанному ниже.

В то время, как другие современные методы едва справились с переходом наклона 2 (SIFT), 2,5 (Harris-Affine и Hessian-Affine), ASIFT может обрабатывать переходы наклона до 32 и выше. MSER может фактически работать с переходом наклона до 10 только тогда, когда оба объекта находятся примерно на одном расстоянии. В самом деле, в отличие от SIFT, MSER не масштабно-инвариантен, потому что он не имитирует размытие из-за увеличения расстояния до объекта. Как показали эксперименты с изображениями, на которых большинство сцен с незначительным или умеренным изменением угла обзора камеры, результаты ASIFT совпадают с SIFT (с обычно

меньшим числом точек совпадения). Но, когда изменение угла обзора становится важным, SIFT и другие методы не срабатывают в то время как ASIFT продолжает работать.

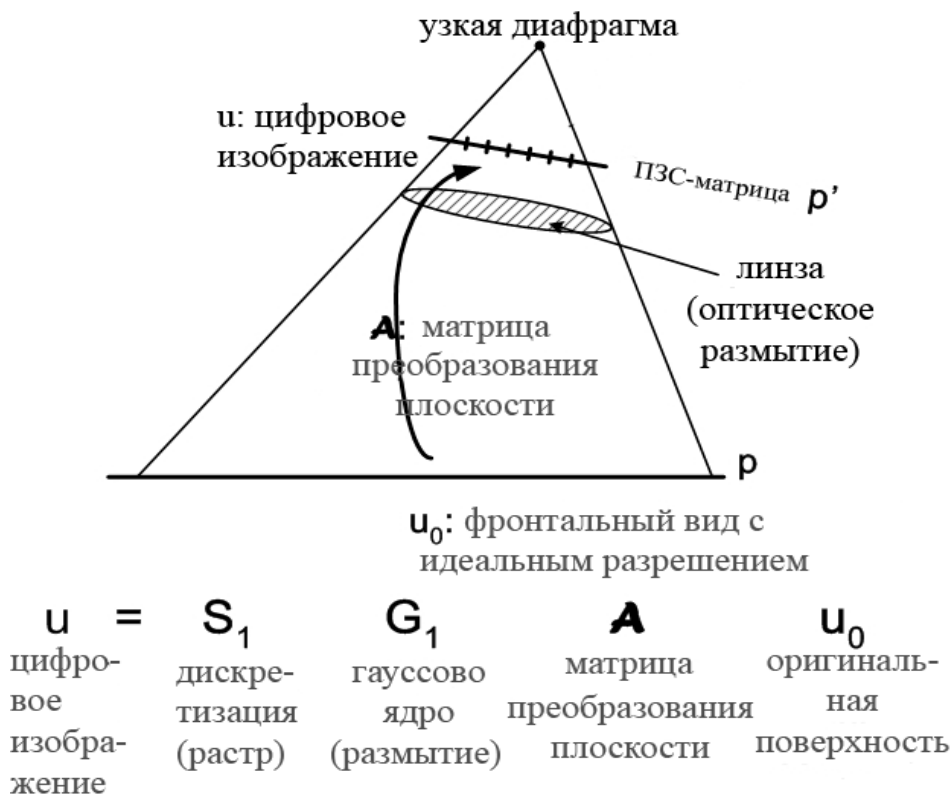


Рисунок 2 – Проекционная модель камеры.

Как видно на примере модели камеры на рисунке 2 получение цифрового изображения плоского объекта может быть описано как

$$u = S_1 G_1 A T u_0,$$

где u - цифровое изображение, а u_0 - фронтальный вид плоского объекта с бесконечным(идеальным) разрешением. Параметры T и A , соответственно, смещение плоскости и матрица преобразования плоскости из-за перемещения камеры. G_1 - гауссова свертка, моделирующая оптическое размытие, а S_1 - стандартный оператор дискретизации на регулярной решётке с плотностью ячеек 1. Гауссовское ядро берётся достаточно широким для того, чтобы убедиться в отсутствии сглаживания по 1 ячейке, поэтому с интерполяцией Шеннона-Уиттакера I , можно восстановить непрерывное изображение из его дискретной версии: $IS_1 G_1 A T u_0 = G_1 A T u_0$. Таким образом в дальнейшем мы будем пренебрегать S_1 .

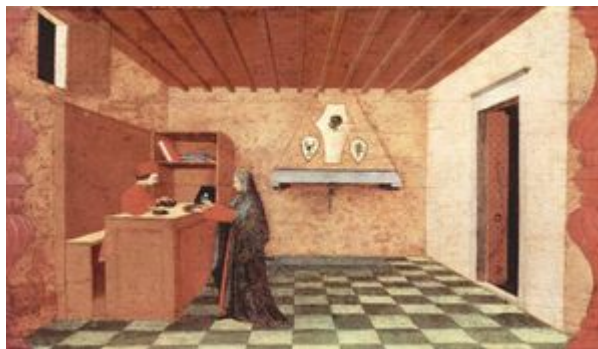


Рисунок 3 – Локальная аффинная аппроксимация.

Перейдем к дальнейшему упрощению изложенной выше модели за счет приближения A матрицей аффинного преобразования. На рисунке 3 показана одна из первых перспективно правильных картин Паоло Уччелло времён эпохи Ренессанса. На полу заметно сильное перспективное искажение: прямоугольные плитки становятся трапециями. Тем не менее, каждая плитка на тротуаре почти параллелограмм. Это иллюстрирует локальное совпадение перспективных деформаций с аффинными преобразованиями. Действительно, в первом приближении по формуле Тейлора любая плоская гладкая деформация может быть приближённо описана вокруг каждой точки аффинным отображением. Перспективные деформации плоского объекта, вызванные движением камеры являются плоским гомографическим преобразованием, которое является гладким, и, следовательно, локально совпадающим с аффинным преобразованием $u(x, y) \rightarrow u(ax + by + e, cx + dy + e)$ в каждой области изображения.

Любое аффинное отображение A со строго положительным определителем имеет единственное разложение

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} = H_\lambda R_1(\psi) T_t R_2(\phi) = \lambda \begin{bmatrix} \cos \psi & -\sin \psi \\ \sin \psi & \cos \psi \end{bmatrix} \begin{bmatrix} t & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \cos \phi & -\sin \phi \\ \sin \phi & \cos \phi \end{bmatrix} \text{ где } \lambda > 0, \lambda t$$

– определитель матрицы A , R_i – повороты, $\phi \in [0, \pi)$, а T_t – наклон, а именно диагональная матрица с первым собственным значением $t > 1$, а вторым равным 1 .

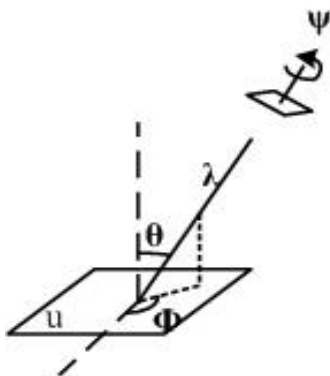


Рисунок 4 – Геометрическая интерпретация аффинного разложения.

На рисунке 4 показана интерпретация движения камеры аффинным разложением: φ и $\theta = \arccos 1/t$ являются углами точки обзора, ψ параметризует вращение камеры вокруг своей оси и λ соответствует зуму. Предполагается, что камера (малый параллелограмм верху справа) находится далеко от изображения u и в начальном положении направлена фронтально, т. е. $\lambda = 1$, $t = 1$, $\varphi = \psi = 0$. Во-первых, камера может двигаться параллельно плоскости объекта: это движение вызывает смещение T , что устраняется, если предположить без ограничения общности, что ось камеры пересекает плоскость изображения в фиксированной точке. Плоскость, содержащая нормальную и оптическую ось составляет угол φ с фиксированной вертикальной плоскостью. Этот угол называется "долгота". Его оптическая ось составляет угол θ с нормалью к плоскости u изображения. Этот параметр называется "широта". Оба параметра являются классическими координатами на «полушарии наблюдения». Камера может вращаться вокруг своей оптической оси (параметр вращения ψ). И наконец камера может перемещаться вперед и назад, что измеряется параметром λ - зумом. Мы убедились, что аффинная модель достаточно хороша, чтобы описать локальные проективные деформации. Если камера не считается расположенной достаточно далеко, деформации плоскости изображения при движении камеры будут гомографией. Тем не менее, как указывалось выше, гомография локально совпадает с аффинным отображением.

Параметр t , определенный выше, называется *абсолютным наклоном*, так как он измеряет наклон между *фронтальным* видом и *наклонным* видом. В реальных приложениях, оба сравниваемых изображения, как правило, видны под наклоном. *Переход наклона* предназначен для определения величины наклона между двумя такими изображениями. Предположим, что $v(x,y) = u(A(x,y))$ и $w(x,y) = u(B(x,y))$ - два наклонных вида плоской сцены, образ которой $u(x,y)$, где A и B являются матрицами двух аффинных преобразований. Тогда $v(x,y) = w(AB^{-1}(x,y))$. Переход наклона между V и W определяется как абсолютный наклон связанный с аффинным отображением AB^{-1} . Пусть t и t' абсолютные наклоны двух изображений u и u' , и пусть φ и φ' - их углы долготы. *Переход наклона* $\tau(u,u')$ между двумя изображениями зависит от абсолютного наклона и углов долготы, и удовлетворяет условию

$$t/t' \leq \tau(u, u') = \tau(u', u) \leq t t'$$

где мы предполагаем, что $t \geq t'$. Переход наклона может быть гораздо выше, чем абсолютный наклон. Следовательно, важно, чтобы алгоритм сопоставления изображений должны был инвариантен к высоким переходам наклона.

Рисунок 5 иллюстрирует пример высокого перехода наклона. Фронтальное изображение (вверху) сжимается в одном направлении на картинке слева при одной точке обзора, и в перпендикулярном направлении при другой точке обзора на картинке справа. *Абсолютный*

наклон (коэффициент сжатия) составляет около 6 в каждом представлении. В результате коэффициент сжатия, или *переход наклона* слева направо, на самом деле $6 \times 6 = 36$.

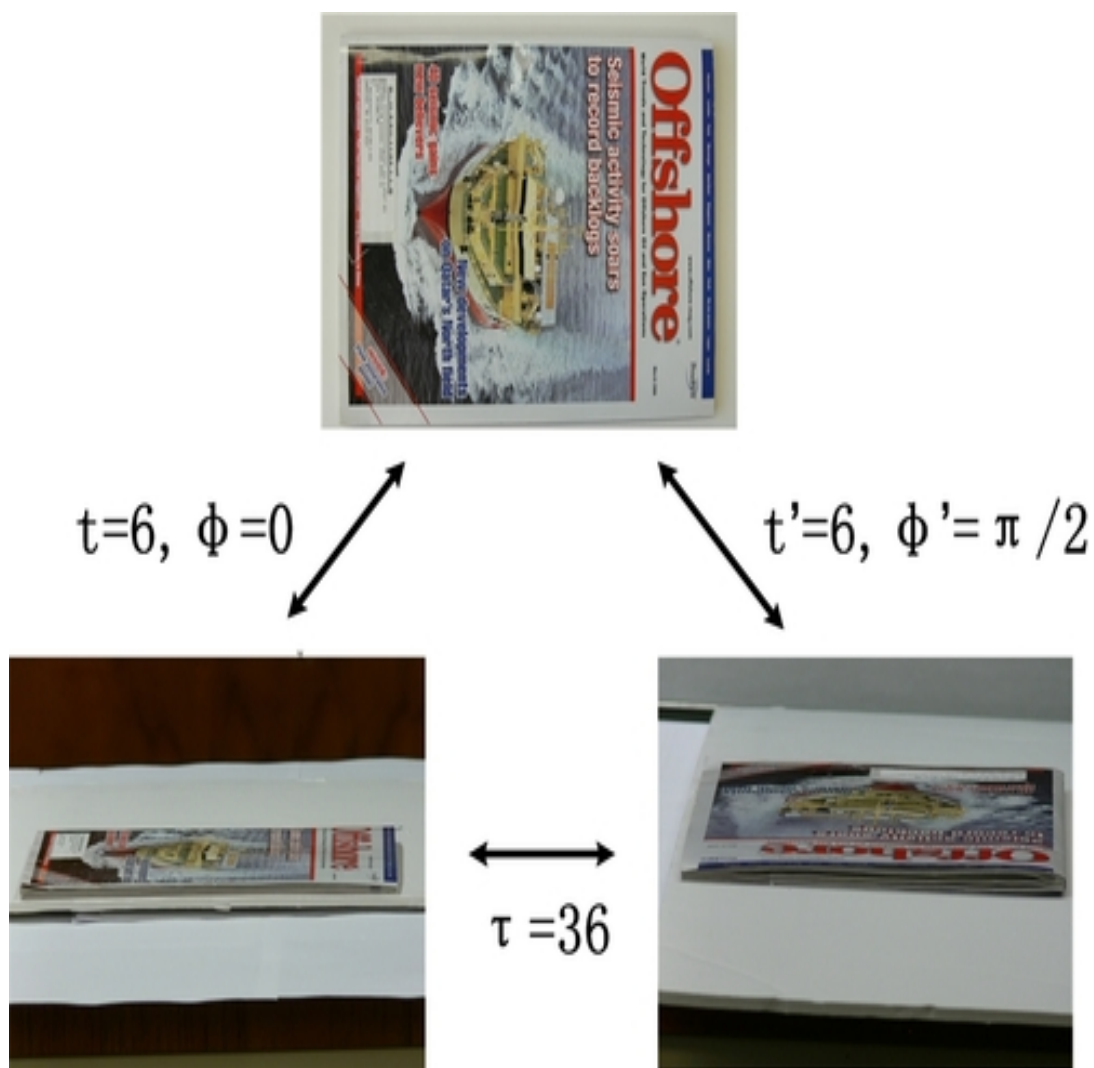


Рисунок 5 – Высокий переход наклона.

Полностью аффинно-инвариантный алгоритм сопоставления изображений должен охватывать все шесть аффинных параметров. Метод SIFT охватывает 4 параметра путем нормализации *вращений и смещений* и моделирования всех *масштабов* от изображения-запроса и изображений для поиска.

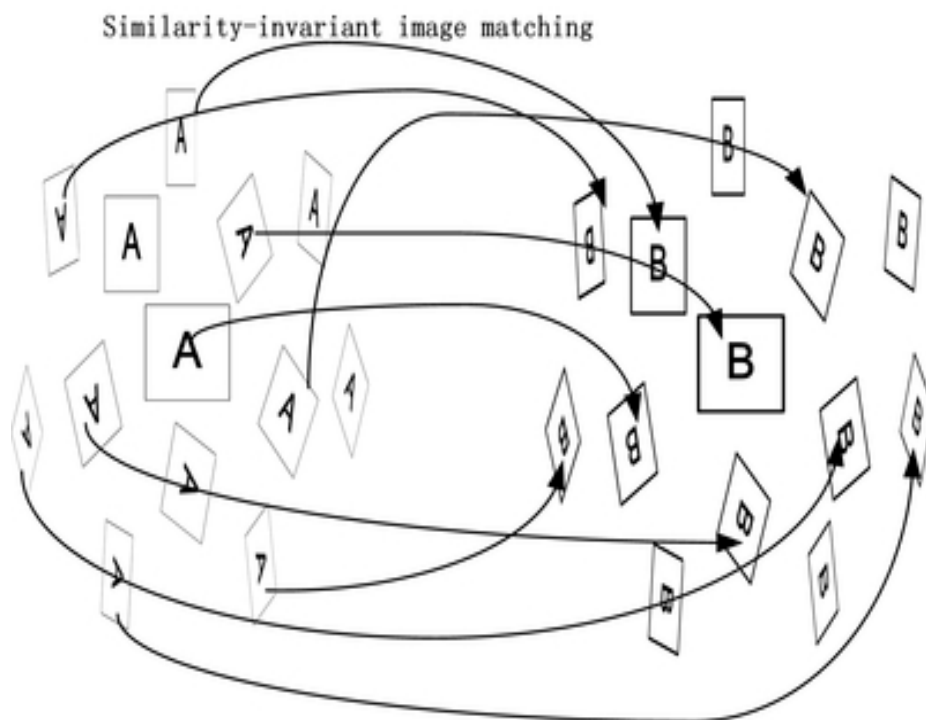


Рисунок 6 – Алгоритм ASIFT.

Как показано на рисунке 1.3.9, ASIFT дополняет SIFT, имитируя два параметра, моделирующих направление оптической оси камеры (оригинал и смоделированные изображения представлены, соответственно, квадратом и параллелограммами), а затем применяет метод SIFT чтобы сравнить смоделированные изображения, так что все 6 параметров оказываются покрыты.

Другими словами, ASIFT имитирует три параметра: масштаб, углы наклона камеры – *долготу* и *широту* (что эквивалентно *наклону*) и нормализует остальные три (*смещение* и *вращение*). ASIFT, таким образом, полностью аффинно-инвариантен.

Выполнение алгоритма ASIFT состоит из следующих этапов:

1. Каждое изображение преобразуется путем имитации всех возможных аффинных искажений, вызванных изменением оптической ориентации оси камеры относительно фронтальной позиции. Эти искажения зависят от двух параметров: долготы φ и широты θ . Изображения поворачиваются на угол φ после чего следуют наклоны с параметром $t = 1/|\cos \theta|$ (наклон t в направлении x является операцией $u(x, y) \rightarrow U(t x, y)$). Для цифровых изображений, наклон осуществляется созданием t -подвыборки. Поэтому она требует предварительного применения сглаживающего фильтра в направлении x , а именно свертки гауссианом со стандартным отклонением $c\sqrt{t^2 - 1}$. Эти повороты и наклоны выполняются для конечного и небольшого количества углов широты и долготы. Шаги этих параметров

обеспечивают нахождение смоделированных изображений вблизи любого другого возможного вида, порожденного другими значениями φ и θ (см. ниже).

2. Все смоделированные изображения сравниваются алгоритмом SIFT. SIFT можно заменить любым другим инвариантным алгоритмом поиска сходств. (Есть много таких вариантов SIFT).
3. Алгоритм SIFT имеет свой собственный критерий отбора неправильных совпадений. Тем не менее, он обычно пропускает много ложных совпадений, даже в парах изображений, которые не соответствуют одной и той же сцене. ASIFT, сравнивая многие пары, поэтому может накапливать много неправильных совпадений. Важно отфильтровать эти совпадения. Критерием является то, что сохранённые совпадения должны быть совместимы с эпиполярной геометрией. Мы используем для этой цели метод ORSA, который считается самым надежным методом, более устойчивым к ошибкам, чем классическая процедура RANSAC. Таким образом, может случиться, что два изображения не имеют совпадений. Это не обязательно означает, что нет ASIFT совпадений; все совпадения могут быть устранены как несовместимые с эпиполярной геометрией.

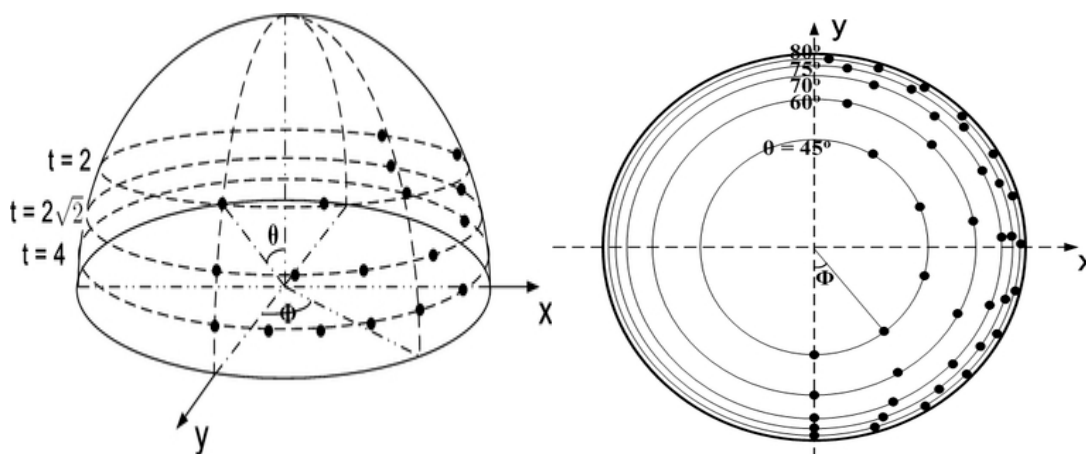


Рисунок 7 – Отбор параметров. Образцы обозначены черными точками. Слева – фронтальный вид на полушарие наблюдений, справа – вид сверху.

Точность выбора параметров широты и долготы должна возрастать с θ , так как искажения изображения, вызванные смещением угла широты или долготы, более существенны при больших θ . Как показано на рисунке 6, отбор пар параметров широты и долготы производится следующим образом:

- Широты θ отбираются таким образом, чтобы связанные с ними наклоны следовали в геометрической прогрессии $1, a, a^2, \dots, a^n$, с $a > 1$. Выбор $a = \sqrt{2}$ является хорошим компромиссом между точностью и шириной шага. В текущей реализации,

значения n находятся в пределах до 5. В результате можно исследовать переходы наклона, возрастающие до 32 (и даже немного больше).

- Долготы φ для каждого наклона являются арифметической прогрессией $0, b / t, \dots, kb / t$, где $b \simeq 72^\circ$, кажется, является хорошим компромиссом, а k это максимальное целое число такое, что $kb / t < 180^\circ$.

Обработка информации, полученной из анализа точек интереса

Общее (полученное по всем изображениям) многомерное облако векторов, характеризующих точки интереса, подвергается PCA преобразованию [3] – отображается в 119-мерное пространство главных компонент – с последующей его сферизацией (в другой терминологии – отбеливанием). Высокая размерность нового признакового пространства обусловлена высокой информационной ёмкостью исходных SIFT-дескрипторов. Сферизация облака (нормировка данных на единичную дисперсию) необходима вследствие задания в новом пространстве евклидовой метрики.

В результате каждое изображение из набора входных изображений представляет собой множество 119-мерных векторов. Для отображения этих векторов в визуальные слова необходимо создать словарь этих слов. Словарь создаётся посредством кластеризации (сферизованного) многомерного облака. Результатом кластеризации является набор кластеров, каждый из которых и есть искомое визуальное слово [1]. Кластер обычно (хотя и не обязательно) описывают его центром масс. В данной работе мы придерживаемся именно такого подхода. Под отображением вектора в визуальное слово понимается его замена ближайшим (по евклидовой метрике) визуальным словом.

В ходе работы над проектом проводились эксперименты с двумя наиболее популярными алгоритмами кластеризации – алгоритмом К-средних и алгоритмом кластеризации на основе самоорганизующихся карт Кохонена (SOM-кластеризация). Согласно экспериментальным данным, алгоритм К-средних не даёт требуемого для решения поставленной задачи разбиения многомерного облака, которое может быть достигнуто только иерархической (снизу-вверх) SOM-кластеризацией. Достаточный уровень иерархии – 3.

Основными критериями разбиения является взаимное расположение точек в многомерном пространстве – требуется, чтобы расстояние между выделенными группами точек было как можно большим, в то время как расстояние между точками в каждой группе должно быть минимальным.

Дескриптор каждого изображения вычисляется, как среднее (по сумке) всех визуальных слов, взвешенных с tf-idf коэффициентами. Вес tf-idf есть произведение двух членов – tf (term frequency) и idf (inverse document frequency). Для произвольного визуального слова член tf есть частота появления данного слова на изображении; член df есть частота появления изображения,

содержащего данное слово, в множестве всех рассматриваемых изображений; член idf есть логарифм величины, обратной к df . Если визуальное слово “характерно” для изображения – оно получает больший tf-idf вес, если оно является “типичным” – меньший.

На рисунке 8 представлен алгоритм работы системы в режиме обучения.



Рисунок 8 – Алгоритм работы системы в режиме обучения.

Локализация БЛА (оценка его географических координат) осуществляется на основе текущего изображения подстилающей поверхности (изображения-запроса) и рабочей БД изображений. Алгоритм работы системы состоит из следующих этапов:

- Вычисление дескриптора изображения-запроса.
- Расчет оценки степени сходства двух дескрипторов посредством нормированного скалярного произведения. Выбор данной “метрики” обусловлен возможностью достаточно быстрого её вычисления – при объёме рабочей БД вплоть до 10^6 изображений, поиск наиболее схожего изображения запрашиваемому может происходить в режиме реального времени [4].
- Находятся соответствующие точки (интереса) на изображении-запросе и изображении-результате распознавания, становится возможным оценить текущие географические координаты БЛА, так как все изображения рабочей БД геопривязаны.

Выводы и область применения

В ходе работы был проведен анализ современного состояния и тенденций развития методов и алгоритмов анализа и оценки видеоданных для решения схожих задач. Проанализированы известные алгоритмы, позволяющие распознавать и выбирать из базы данных изображений наиболее похожие на предъявленное, реализованы новые алгоритмы, позволяющие повысить эффективность отдельных операций технологического цикла обработки и интерпретации изображений. Полученные результаты показали возможность применения подобной системы для решения задач локализации БЛА, однако алгоритмы нуждаются в дальнейшей доработке с целью повышения их эффективности. Одной из задач дальнейших исследований является совершенствование алгоритмов поиска характерных точек на изображениях подстилающей поверхности для достижения большей устойчивости системы к шумам и искажениям.

Библиографический список

1. Lutsiv V.R., Malyshev I.A., Pepelka V., Potapov A.S. Target independent algorithms for description and structural matching of aerospace photographs // Proc. SPIE. 2002. V. 4741. P. 351–362
2. Popescu A., Moëllic P.-A. An ontology driven image retrieval system. // Proc. 6th ACM Int. conf. on Image and Video Retrieval. 2007. P. 113–116
3. Hörster E., Lienhart R., Slaney M. Image retrieval on large-scale image databases // Proc. 6th ACM int. conf. on Image and Video Retrieval, CIVR 2007. 2007. P. 17–24.
4. Rares A., Reinders M.J.T., Hendriks E.A. Image interpretation systems // Technical Report (MCCWS 2.1.1.3.C), MCCWS project, Information and Communication Theory Group. TU Delft. 1999. P. 32.

Сведения об авторах

ПУРТОВ Илья Сергеевич, научный сотрудник Института истории, естествознания и техники,
тел.: +7 (916) 230-66-88 e-mail: ilya.purtov@gmail.com.

СИНЧА Дмитрий Петрович,
тел.: + 7 (499) 504-14-97, e-mail: atsglance@list.ru