Санкт-Петербургский политехнический университет Петра Великого

Институт машиностроения, материалов и транспорта

Высшая школа автоматизации и робототехники

ОТЧЁТ ПО БИЛЕТУ

По дисциплине: Системы технического зрения

Выполнил А.М.Орехов  
студент группы №3331506/00401

Принял В.В.Титов

«\_\_\_» \_\_\_\_\_\_\_\_\_\_\_\_\_2024 г

Санкт-Петербург

2024

# Билет №5. Оценка аффинных преобразований и гомографии.

Гомографией называют проективное преобразование одной плоскости в другую. В случае компьютерного зрения, за плоскости принимают изображения или плоские поверхности в трехмерном пространстве. К применениям гомографии относят:

* Ректификация (Выпрямление) изображений
* Сшивание панорам из отдельных изображений
* Деформация текстур в трехмерной графике
* Определение положения камеры
* Многовидовая геометрия

По существу, гомография H отображает точки плоскости (в однородных координатах) согласно следующему уравнению:

|  |  |  |
| --- | --- | --- |
|  | или |  |

Однородные координаты – система координат, использующаяся в теме проективной геометрии. Причиной этому является свойство, гарантирующее что точка, представленная в однородных координатах, при умножении на ненулевой масштабный коэффициент, не изменяется. Также, зачастую производится нормировка: полагают, что и с помощью этого получается однозначно представлять точки изображения только координатами x и y.

Таким образом, получается, что гомография H определена с точностью до масштабного коэффициента w и имеет восемь независимых степеней свободы.

Физический смысл параметров матрицы проективного преобразования следующий:

* Параметры a, b, d, e являются матрицей поворота, а дополнительное домножение параметров a и e, находящихся на главной диагонали матрицы дополнительно масштабирует точки.
* Параметры c и f задают параллельный перенос
* Параметры g и h задают изменение перспективы

Параметры проективного преобразования между двумя изображениями можно определить непосредственно, зная соответственные точки двух изображений. В общем случае у проективного преобразования 8 степеней свободы. Каждая найденная пара соответствующих точек дает возможность составить два уравнения. Соответственно, в минимальном варианте для оценки гомографии требуется всего 4 пары точек. Впрочем, для большей стойкости алгоритма к возможным неточностям определения точек, использование минимально возможного количества нежелательно.

Переписав матричную операцию, можно получить уравнения для компонент x и y:

Запишем уравнения в следующей форме:

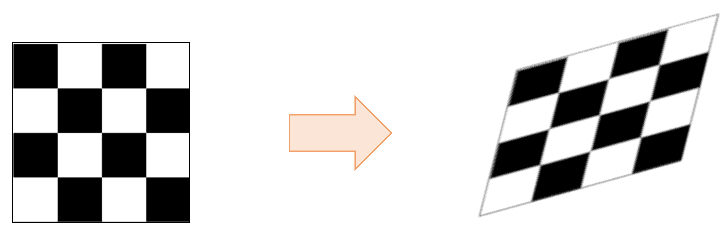
Где i – индекс пары соответствующих точек. Теперь, если переписать уравнение проективного преобразования, подставив в него полученные уравнения, можно получить матрицу вида:

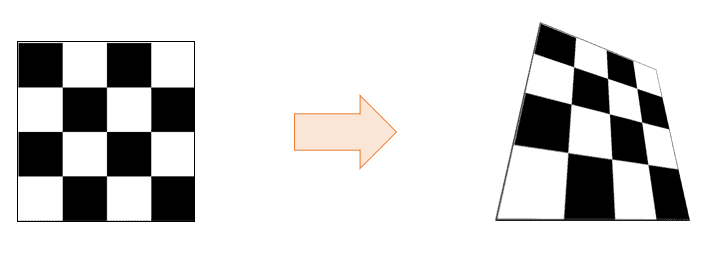
Уравнение подобного вида является основой метода решения путем прямого линейного преобразования. Далее по алгоритму из этого уравнения потребуется найти нетривиальное значение h – векторизованной матрицы H:

Кроме гомографии, также существует несколько важных частных случаев проективных преобразований. Одним из таких является ***аффинное преобразование***.

|  |  |  |
| --- | --- | --- |
|  | или |  |

Данный вид преобразований отличается от гомографии отсутствием параметров перспективных изменений, но сохраняют преобразования поворота, масштабирования и параллельного переноса. При аффинном преобразовании, на двух изображениях сохраняется параллельность прямых. В случае проективного преобразования сохраняется только прямолинейность отдельных линий (Рисунок 1).





*Рисунок 1 – Аффинное и проективное преобразование*

Для вычисления параметров аффинного преобразования можно применить метод прямого линейного преобразования, приравняв последние два элемента к нулю () Аналогично гомографии, поскольку у аффинного преобразования только шесть степеней свободы, для решения уравнения потребуется уже три пары соответствующих точек.

Таким образом, были рассмотрены основы прямого линейного преобразования для гомографии и аффинного преобразования. Была получена система уравнений в матричной форме, которая описывает соответствие между соответствующими точками на исходном и преобразованном изображении. Следующим шагом является нахождение вектора параметров ***h*** и его последующее преобразование обратно в матрицу **H**. Для идеального случая с отсутствием шумов и идеально выбранным парам соответствующих точек возможно напрямую найти значение матрицы. Однако, зачастую составленная система будет иметь количество уравнений больше, чем число неизвестных, а, следовательно, не существует такого вектора h, который бы являлся решением сразу для всех уравнений.

В реальных условиях нахождение решения только по четырем парам точек может быть усложнено из-за шумов, ошибок измерения, выбросов и ошибочно сопоставленных точек, не являющихся на самом деле соответствующими. Для решения этой проблемы требуется выбирать количество соответствующих точек большее, чем требуется для выбранной модели преобразования, улучшая результат за счет избыточности. В таком случае, для нахождения решения уравнения потребуется применение методов, предполагающих нахождение оптимального решения, пользуясь определенной метрикой, например, метод наименьших квадратов. Он, впрочем, не всегда может давать корректные результаты, в случаях, когда кроме «правильных» точек в набор были выбраны и ложные соответствия. Из-за подобной особенности как правило применяются робастные методы, позволяющие не учитывать шумы и ложные соответствия, например, RANSAC.

# Билет №33. Эпохи решения задачи классификации.

Задача классификации является одной из ключевых проблем в области машинного зрения, нацеленной на разделение объектов, сцен, или событий на определенные категории на основе формализованных визуальных признаков. На протяжении нескольких десятилетий эта область исследований пережила несколько эпох, каждая из которых отмечена появлением новых методов, технологий и подходов, улучшавших точность и эффективность классификации.

В начале 1960-х компьютерное зрение зарождалось в университетах, изучающих вопросы искусственного интеллекта и одними из первых решаемых в этой сфере задач, была задача классификации геометрических объектов. Одним из препятствий для создания алгоритма классификации объектов является непосредственно разнообразие различных классов объектов. Объекты могут иметь различные формы, размеры, текстуры и цвета, что делает сложной задачу их точного выделения и классификации. На заре сферы компьютерного зрения, данную проблема решалась уменьшением разнообразия классов для распознавания и упрощением их признаков. Так, в 1963 году вышла диссертационная работа, посвященная классификации трехмерных объектов на изображении.

Изображение выглядит как текст, диаграмма, зарисовка, дизайн

Автоматически созданное описание

Рисунок 1 - Определение трехмерных тел. L.G.Roberrts Ph.D thesis

Примененный метод на полученном изображении выделял острые углы, находя их с помощью дифференциального оператора, после чего применялся многоэтапный процесс по выделению линий. Первым шагом по изображению проходило окно 4х4, находившее точки, потенциально принадлежащие линиям. После чего над изображением проводилось 4 операции корреляции с целью найти линии со строго определенными направлениями – 0, 1, и -1. Далее, если прямые соприкасались друг с другом в окне пикселей 4х4, они соединялись и на место соединения добавлялась точка. Все линии без соединений удалялись. На собранных формах выполнялись операции по поиску фигур, сначала собирая линии в двумерные примитивы, а после сопоставляя их форму с формами трехмерных примитивов.

Позже, в 1966 группа ученых из MIT для работы над задачей инициировали «summer vision project» как эксперимент по созданию программы, которая позволит компьютеру «объяснять то, что он видит». Первичной целью, сформулированной авторами, было разделить область кадра на объекты, фон и «хаос». В качестве плана-максимум планировалось решить задачу идентификации объектов в кадре по предварительно известной библиотеке объектов, в качестве которых были выбраны геометрические примитивы – шары, кубики с одинаковыми и разными цветами граней и цилиндры.

Уже на данном этапе становится видна тенденция, что разрабатываемые алгоритмы стараются рассмотреть объекты как совокупность более простых составляющих. Отдельно стоят рассмотрения объемные примитивы, поскольку они позволяют разбить задачу классификации. Ограниченный набор примитивов будет иметь предсказуемые проекции на экране. При этом, отсекаются несущественные детали поверхности, после чего, основной объект классификации можно оценить по соотношению примитивов. Кроме того, подобный метод позволяет уменьшить затрат памяти программы, поскольку общее количество элементарных составных частей будет всегда меньше, чем число объектов, которые можно из них составить. Это было особенно важно на ранних этапах развития систем технического зрения по причине ограниченности вычислительных ресурсов того времени.

Важной работой, реализующей этот принцип являются «обобщенные цилиндры», предложенные Бинфордом в неопубликованной работе 1971 года. Обобщенный цилиндр – это тело, образованное одномерным набором секущих поверхностей, переходящих одна в другую плавной деформацией. Данное определение отражает следующую идею: многие интересные объекты при надлежащем уровне детализации сводятся к поверхностям подобного рода.

В связи с достаточно общим определением обобщенного цилиндра, нельзя сформировать модель, достоверно дающую ответ на вопрос как обобщенный цилиндр будет выглядеть на изображении. Таким образом, у нас не может быть точных геометрических соображений как были использованы в прошлом рассмотренном методе. Однако, возможно найти на изображении формы, которые выглядят так, как будто они порождены методом переноса эскиза сечения по определенной траектории (заметания). В данном методе подобные фигуры были названы лентами.

Лента – это огибающая семейства копий геометрической фигуры (образующей ленты), движущейся по некоторой траектории (хребту ленты), сжимающейся или расширяющейся по мере движения. Важные частные случаи лент – лента Брукса получается, если образующая это фрагмент прямой, а если образующей является окружность, получается лента Блюма. На рисунке ниже показаны объекты, нарисованные алгоритмом ACRONYM, основанным на описанном принципе для самолетов Boeing 747 и Lockheed L-1011:

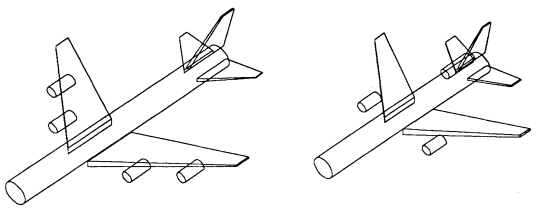
 

Рисунок 2 - Определение структуры самолета методом ACRONYM

Можно видеть, как алгоритм распознал не только общую структуру самолета, но и отдельные составляющие элементы вроде двигателей, что в теории позволяет распознавать в том числе разновидности внутри класса, решая проблему внутриклассового разнообразия.

Однако, несмотря на то что общая идея сформулирована, еще нет рабочей программы, которую можно было бы назвать по-настоящему универсальной, поскольку реализация описанной идеи требует решения дополнительных практических задач. Например, выбор принципа группировки и организации представлений лент в объекты.

Другой принцип был предложен Ирвином Бидерманом в 1987. Идея использования фиксированный набор трехмерных объектов, названных геомами. Всего было выделенно 36 фигур, определены правила для распознавания их проекций на изображениях.

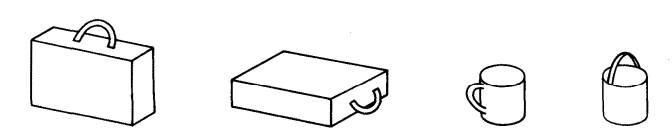
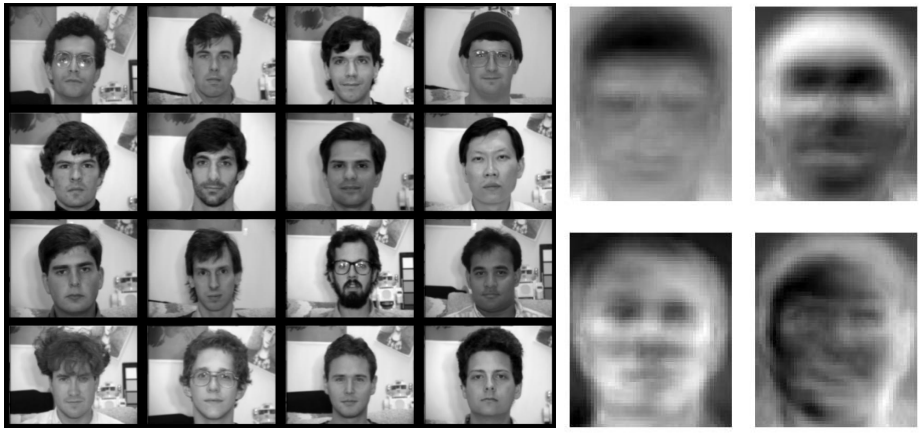


Рисунок 3 – Различные конфигурации геомов составляют разные объекты

Однако, можно заметить, что рассматриваемые в ранних работах методы, основываясь на геометрических примитивах плохо применяются для сложных объектов реального мира. В том числе, важной задачей из сферы классификации является распознавание лиц. В отличие от описанных ранее задач, структура лица сохраняется у всех людей, а для классификации требуется принимать во внимание детали, которые мы отбрасывали в описанных ранее методах для геометрических фигур.

Важной работой, направленной на решение данной проблемы для задачи с лицами стала «Eigenfaces for Recognition» Мэтью Турка и Алекса Пентланда (Matthew A. Turk & Alex P. Pentland), предложивший для распознавания лиц использовать лишь их изображения, что относится к группе методов, основанных на внешности (appearance-based methods). Примененный подход использует так называемые «собственные лица» (eigenfaces), названные по аналогии с собственными числами и векторами матрицы (eigen values, eigen vectors). Последовательность работы алгоритма следующая:

1. Инициализация. Необходимо собрать определенное количество изображений с разными выражениями лица, освещением и ракурсами. Все изображения должны иметь одинаковый размер (в оригинальной статье используется 256 на 256) и выровнены по глазам.
2. Векторизация изображений – Каждое изображение лица преобразуется в одномерный вектор путем конкатенации строк в одну последовательность. Это позволяет работать с изображениями как с векторами в высокоразмерном пространстве.
3. Вычисление матрицы ковариации: на основе всех векторов-изображений вычисляется ковариационная матрица.
4. Для найденной матрицы находятся собственные вектора и значения. Собственные векторы, соответствующие наибольшим собственным значениями, и являются eigenfaces.
5. Формируется M-мерное (где M – число изображений в выборке) подпространство eigenfaces: если изначально векторизованное изображение представляло собой вектор в пространстве размерностью 256x256, однако лишь часть этого пространства соответствовала изображениям лиц. Описанным методом выбирается определенное количество векторов, которые формируют новую систему координат в пространстве изображений, называемую «пространством лиц» (face space).
6. Для классификации новые изображения проецируются на пространство eigenfaces и вычисляются его координаты в этом пространстве. Эти координаты представляют собой весовые коэффициенты, отражающие вклад каждого лица из исходной выборки в новое изображение. На основании схожести координат лицо относится к тому или иному человеку из первоначальной выборки.



• Mid-late 1990s: sliding window approaches

• Late 1990s: feature-based methods

• Early 2000s: parts-and-shape models

• 2000 – 2010: bags of features

• 2010-2015: combination of local and global methods, modeling context, integrating recognition and segmentation

• 2015 – present: machine learning, artificial neural network