笔记本：	LiteratureReview		
创建时间：	2020/6/19 10:20	更新时间：	2020/6/29 9:27
作者：	haokun-wang@outlook.com		

AprilTag: A robust and flexible visual fiducial system

Edwin Olson

Abstruct

Aritificial feaatures play an important role in systems where *perception* **is not** the central objective.
AprilTag is a new visual fiducial system that uses a **2D bar code** style "tag", allowing full **6 DOF** localization of features from **a single image**.
This method is fully open and the algorithms are documented in detail.

利用人工设计的tag，从单张2D图片中估计出6D的姿态，相比于已有的ARTag等，有开源、文档丰富的优点。

Introduction

Although related to other 2D barcode like QR codes, Apriltag has significantly goals and applications.
QR code usually involved in **aligning the camera** with the tag, and photographs it at fairly high resolution **obtaining hundreds of bytes**.
Apriltag has a small information payload (perhaps 12 bits), but is designed to be **automatically detected and localized**.
Fiducial systems also are designed to detect **multiple markers** in a **single image**.

ARToolkit and **ARTag** are the best known for the application to augmented reality.
Designing robust fiducials is a challenging both from **a marker detection standpoint** (which pixels in the image correspond to a tag?) and from **a error-tolerrant data coding standpoint** (which tag is it?)
Contributions:

1. Propose a graph-based image **segmentation algorithm** based on local gradients that allows lines to be precisely estimated, and describe **a quad extraction method** that can handle significant occlusions.
2. Describe **a new coding system** that address problems unique to 2D barcoding systems: robustness to rotation, and robustness to false positives arising from natural imagery.
3. Provide results on **a set of benchmarks**.

二维码主要是用来标定相机，或者在高分辨率画面中提供数百bytes的信息，比如网址等。AprilTag是用来被自动识别和定位的，因此只需包含更小的信息载荷。

设计一款鲁棒性强的基准系统主要有两个挑战：一是从2D图形中识别出属于标记物的像素；二是提高标记物的容错机制，使得不同标记物能被清晰地区分开。

文章的主要贡献包括：一提出了一个基于图理论的图片分割算法，描述了一种能处理显著遮挡情况的四边形特征提取方法；二描述了一个新的编码系统，解决旋转、正样本误检的鲁棒性；三将测试结果做成了基准数据集。

Previous Work

ARToolkit was targeted at artificial reality applications. Its tags contained a square-shaped payload surrounded by a black border. It was encoded in symbols such as the latin characters. When decoding a tag, the payload of the tag was correlated against a database of known tags.
A major disadvantage of this approach is the computational cost associated with decoding tags, since each template requireda separate, slow correlation operation. A second disadvantage is that it is difficult to generate templates that are approximately orthogonal to each other. And the tag detection scheme is based

on a simple binarization of the input image based on a user-specified threshold. This scheme can not handle even modest occlusions of the tag's border.
ARTag provided improved detection and coding schemes. The detection mechanism was based on the image gradient, making it robust to changes in lighting. But the details of the detector algorithm are not public. ARTag provided the first coding system based on *forward error correction*, which made tags easier to generat, faster to correlate, and provided greater orthogonality between tags.
ARToolkitPlus and **Studierstube Tracker** are inspired by ARTag. They introduced *digitally-encoded payloads* like those used in ARTag.
In addition to monochrome tags, other coding systems like encoding color information have been developed. Tags using retro-reflectors have also been used.
Two-dimensional planar systems, like AprilTag, have two advantages over LED-based systems: lower cost, and provide 6 DOF position estimtion without the need of multiple objects.

ARToolkit是一个被黑边包围的正方形信息荷载区域。用符号比如拉丁文编码，识别是通过与已知的tags数据库比对。主要的缺点是识别代价太高，并且很难生成彼此之间有较大区别（正交）的tags. 识别的算子也需要手动指定二值化的阈值，在光照变化的环境下鲁棒性很差。

ARTag在一定程度上弥补了ARToolkit的不足。它的识别机制是基于图片的梯度，因此对光照变化不敏感。但它的识别算法是不开源的。ARTag使用了前向纠错机制，即在生成编码的时候就加入了冗余，以此来提高识别的鲁棒性，并且使得生成正交的tags更加容易。
除了单色的tags，彩色tags也有被提出使用。但是比较而言，二维单色的tags除了生产成本低，也只需要单个tag即提供6D的姿态估算。

Detector

The detector attempts to **find four-sided regions** ("quads") that have a darker interior than exterior.
The quad detector is designed to have a very low false negative rate, and consequently **has a high false positive rate**. We **rely**

on the coding system to reduce this false positive rate to useful levels.

检测器设计的目的是使得负样本出错（属于tag但是被判断成不属于tag，即漏检）的比例非常低，这样会导致正样本出错（不属于tag的部分被判断成属于tag，即误检）的比例升高。我们依赖编码系统来解决正样本出错这个问题。

Detecting line segments

Our approach compute the gradient direction and magnitude at every pixel, and agglomeratively clusters the pixels into components with similar gradient directions and magnitudes. The clustering algorithm is similar to the graph-based method of Felzenszwalb: a graph is created in which each node represents a pixel. Edges are added between adjacent pixels with an edge weight equal to the pixels' difference in gradient direction. Given a component n, we denote the range of gradient directions as D(n) and the range of magnitudes as M(n). Put another way, D(n) and M(n) are scalar values representing the difference between the maximum and minimum values of the gradient direction and magnitude respectively. Given two components n and m, we join them together if both of the conditions below are satisfied:

$$D(n \cup m) \leq \min(D(n),D(m)) + KD/|n \cup m| \qquad (1)$$
$$M(n \cup m) \leq \min(M(n),M(m)) + KM/|n \cup m|$$

Two clusters are joined together if their union is about as uniform as the clusters taken individually. A modest increase in intra-component variation is permitted via the KD and KM parameters, however this rapidly shrinks as the components become larger.

This gradient-based clustering method is sensitive to noise in the image: even modest amounts of noise will cause local gradient directions to vary, inhibiting the growth of the components. The solution to this problem is to low-pass filter the image.

首先是直线检测，这里基于经典的图方法进行了微小改进。对于两个components，如果它们合并后（梯度方向和梯度大小）的变化范围，比之前最小的一个component的变化范围

只多了K，则认为这两个components是可以合并的。缺点是对图像噪声比较敏感，需要一个低通滤波器过滤噪声。

Quad detection

The next task is to find sequences of line segments that form a 4-sided shape, i.e., a quad. The challenge is to do this while being as robust as possible to occlusions and noise in the line segmentations. Robustness to occlusions and segmentation errors is handled by adjusting the “close enough” threshold: by making the threshold large, significant gaps around the edges can be handled.

接着是检测方块，这里的挑战主要来自直线检测的遮挡和噪声。采用了深度优先搜索算法来识别方块。对于所有的直线，选择另一条满足末端足够近且形成逆时针闭环的直线作为搜索树的下一层。其中“足够近”取决于选择的阈值大小，论文中选择了直线长度的2倍＋5个像素。宽松阈值带来的结果就是漏检的比例很小，但是误检的比例比较大。

Homography and extrinsics estimation

即tag坐标系到像素坐标系的变换矩阵求解。

Payload Decoding

We do this by computing the tag-relative coordinates of each bit field, transforming them into image coordinates using the homography, and then thresholding the resulting pixels. In order to be robust to lighting, we use spatially-varying threshold.

We use the border of the tag, which contains known examples of both white and black pixels, to learn this model (see Fig. 4). We use the following intensity model:

$$I(x, y) = Ax + Bxy + Cy + D$$

The threshold used when decoding data bits is then just the average of the predicted intensity values of the black and white models.

解码的步骤包括：从每一个bit在tag坐标系下的位置映射到像素坐标系中，然后经过阈值滤波得到像素的值。为了在不同光照条件下有较好的鲁棒性，我们使用了随空间变化的阈

值计算方法。对黑色块和白色块分别建立预测模型，每个模型需要拟合四个参数，对tag的边界（已知真值）进行采样来拟合。取黑色块和白色块模型在同一个位置预测值的平均作为阈值，来判断这个位置识别的结果是黑色块还是白色块。

Coding System

The goals of a coding system are to:

- Maximize the number of distinguishable codes
- Maximize the number of bit errors that can be detected or corrected
- Minimize the false positive/inter-tag confusion rate
- Minimize the total number of bits per tag (and thus the size of the tag)

These goals are often in conflict, and so a given code represents a trade-off. Classical lexicodes are parameterized by two quantities: the number of bits n in each codeword and the minimum Hamming distance between any two codewords d. Lexicodes can correct $\lfloor (d-1)/2 \rfloor$ bit errors and detect d/2 bit errors. In the case of visual fiducials, the coding scheme must be robust to rotation. In other words, it is critical that when a tag is rotated by 90, 180, or 270 degrees, that it still have a Hamming distance of d from every other code.

编码系统的目标主要有四个：最大化可辨别码的数量；最大化可被纠错或可被检查出错的比特数量；最小化误检比例；最小化每个tag的比特数量。由于几个目标之间是冲突的，通常会在目标间进行妥协。经典的字典编码主要有两个指标：一是每一个编码的比特数n；以及在任意两个编码之间的最小汉明距离d。字典编码可以对$\lfloor (d-1)/2 \rfloor$个字节的错误进行纠正，或检测出d/2个比特的错误。为了满足旋转的鲁棒性，我们也要求tag旋转90, 180或270度时，与其他每一个编码的汉明距离依然不变。另外，为了避免生成容易被误检的简单几何形状（比如一整个黑色块），论文采用了控制生成编码矩形的数量这种方法来避免生成简单几何形状。有复杂模式的tag有较低的误检率在实验中得到了验证。

AprilTag 2:改进了图像识别算法，使用自适应阈值方法分割图像。
AprilTag 3:提出了可变分层的tag结构，比如圆形、环形等；适配了识别算子。