本篇来探讨一下智能手机摄像头中：普通彩色相机（RGB） + 深度相机（Depth）的技术原理。

首先来解释一下什么是深度相机吧。

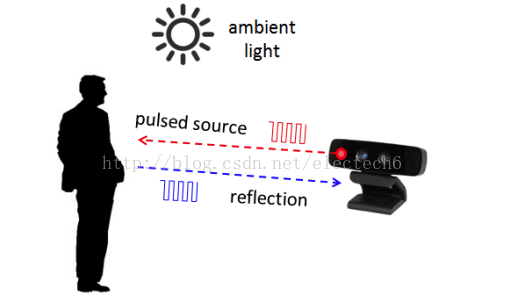
深度相机

顾名思义，深度相机就是可以直接获取场景中物体距离摄像头物理距离的相机。根据原理不同，主要有飞行时间（TOF）、结构光、激光扫描几种方式（注：有些地方将双目立体视觉相机也列为深度相机的一种，它是单纯依靠算法计算得到的深度信息）。目前使用较多的是TOF相机。目前主流的TOF相机厂商有PMD、MESA、Optrima、微软等几家，其中MESA在科研领域使用较多；PMD是唯一在室内/外都能使用的TOF相机厂商，多用于科研、工业等各种场合；Optrima、微软则主要面向家庭、娱乐应用，价位较低。

因为TOF相机使用的较为广泛，在此主要介绍一下TOF相机的原理。

TOF相机

TOF（Time of flight）直译为“飞行时间”。其测距原理是通过给目标连续发送光脉冲，然后用传感器接收从物体返回的光，通过探测光脉冲的飞行（往返）时间来得到目标物距离。这种技术跟3D激光传感器原理基本类似，只不过3D激光传感器是逐点扫描，而TOF相机则是同时得到整幅图像的深度（距离）信息。



TOF组成

TOF相机采用主动光探测，通常包括以下几个部分：

**1、照射单元**

照射单元需要对光源进行脉冲调制之后再进行发射，调制的光脉冲频率可以高达100MHz。因此，在图像拍摄过程中，光源会打开和关闭几千次。各个光脉冲只有几纳秒的时长。相机的曝光时间参数决定了每次成像的脉冲数。

要实现精确测量，必须精确地控制光脉冲，使其具有完全相同的持续时间、上升时间和下降时间。因为即使很小的只是一纳秒的偏差即可产生高达15 c m的距离测量误差。

如此高的调制频率和精度只有采用精良的LED或激光二极管才能实现。

一般照射光源都是采用人眼不可见的红外光源。

**2、光学透镜**

用于汇聚反射光线，在光学传感器上成像。不过与普通光学镜头不同的是这里需要加一个带通滤光片来保证只有与照明光源波长相同的光才能进入。这样做的目的是抑制非相干光源减少噪声，同时防止感光传感器因外部光线干扰而过度曝光。

**3、成像传感器**

是TOF的相机的核心。该传感器结构与普通图像传感器类似，但比图像传感器更复杂，它包含2个或者更多快门，用来在不同时间采样反射光线。因此，TOF芯片像素比一般图像传感器像素尺寸要大得多，一般100um左右。

**4、控制单元**

相机的电子控制单元触发的光脉冲序列与芯片电子快门的开/闭精确同步。它对传感器电荷执行读出和转换，并将它们引导至分析单元和数据接口。

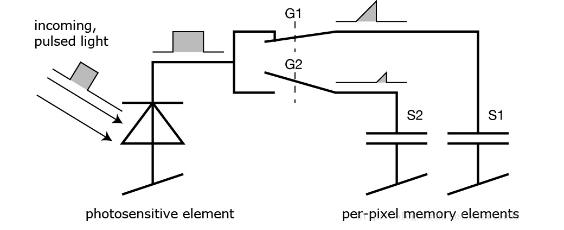
**5、计算单元**

计算单元可以记录精确的深度图。深度图通常是灰度图，其中的每个值代表光反射表面和相机之间的距离。为了得到更好的效果，通常会进行数据校准。

TOF测距原理

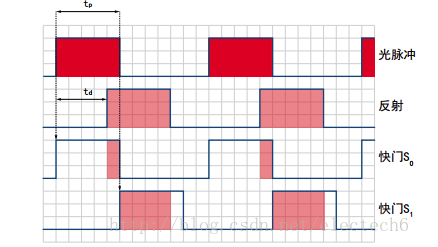
下面来介绍一个经过高度简化的测距原理。

照射光源一般采用方波脉冲调制，这是因为它用数字电路来实现相对容易。深度相机的每个像素都是由一个感光单元（如光电二极管）组成，它可以将入射光转换为电流，感光单元连接着多个高频转换开关（下图的G1，G2）可以把电流导入不同的可以储存电荷(下图S1，S2)的电容里。



相机上的控制单元打开光源然后再关闭，发出一个光脉冲。在同一时刻，控制单元打开和关闭芯片上的电子快门。由光脉冲以这种方式产生的电荷S0存储在感光元件上。

然后，控制单元第二次打开并关闭光源。这次快门打开时间较晚，即在光源被关闭的时间点打开。现在生成的电荷S1也被存储在感光元件上。



因为单个光脉冲的持续时间非常短，此过程会重复几千次，直到达到曝光时间。然后感光传感器中的值会被读出，实际距离可以根据这些值来计算。

记光的速度为c，tp为光脉冲的持续时间， S0表示较早的快门收集的电荷， S1表示延迟的快门收集的电荷，那么距离d可以由如下公式计算：



最小的可测量距离是：在较早的快门期间S0中收集了所有的电荷，而在延迟的快门期间S1没有收集到电荷，即S1 = 0。代入公式会得出最小可测量距离d=0。

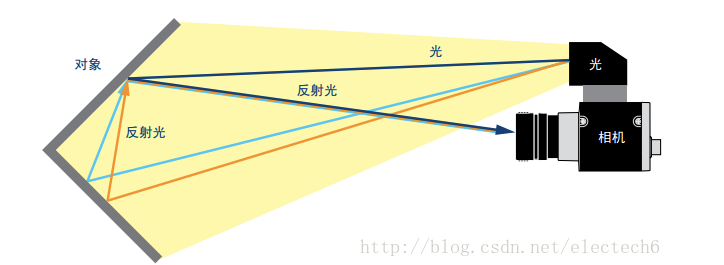
最大的可测量的距离是：在S1中收集了所有电荷，而在S0中根本没有收集到电荷。然后，该公式得出d= 0.5 x c × tp。因此最大可测量距离是通过光脉冲宽度来确定的。例如，tp = 50 ns，代入上式，得到最大测量距离d = 7.5m。

TOF相机影响因素

影响ToF相机的测量精度的因素如下：

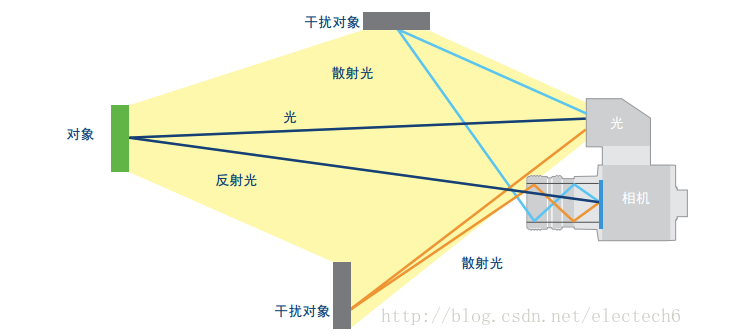
**1、多重反射**。

距离测量要求光只反射一次。但是镜面或者一些角落会导致光线的多次反射，这会导致测量失真。如果多重反射使得光线完全偏转，则没有反射光线进入相机，从而无法正确测量反射面的距离。反之，如果其他方向的光通过镜面反射进入芯片，则可能会发生过度曝光。见下图。



**2、散射光**

在镜头内或在镜头后面发生多余反射会出现散射光，如下图所示，散射光会导致图像褪色，对比度下降等不良影响。所以要避免在相机正前方有强烈反光的物体存在。



**3、环境光**

前面说过，深度相机镜头上会有一个带通滤光片来保证只有与照明光源波长相同的光才能进入，这样可以抑制非相干光源提高信噪比。这种方式确实能够比较有效地过滤掉人造光源，但是，我们常见的日光几乎能够覆盖整个光谱范围，这其中包括和照明光源一样的波长，在某些情况下（如夏天的烈日）这部分光强可以达到很大，会导致感光传感器出现过度曝光。因此相机如果想在这种条件下正常工作，仍然需要额外的保护机制。

**4、温度**

电子元件的精度受温度的影响。所以当温度波动时会影响电子元件的性能，从而影响到脉冲调制的精度。前面说过一纳秒的脉冲偏差即可产生高达15 c m的距离测量误差，因此相机要做好散热，这样才能保证测量精度。

TOF相机输出

那么TOF相机最后输出的是什么呢？

TOF相机内部每个像素经过上述过程都可以得到一个对应的距离，所有的像素点测量的距离就构成了一幅深度图，如下图所示。左边是原图，右边是对应的深度图。



可以看到深度图其实是一张灰度图，它是一张三维的图：水平垂直坐标对应像素位置，该位置的灰度值对应的是该像素距离摄像头的距离。所以深度图中的每个像素可以表示空间中一个点的三维坐标，所以深度图中的每个像素也称为体像素（voxel）。

深度信息融合

当我们获得了深度图后，下一步就是要把深度信息融合到普通RGB相机拍摄的彩色图片。这一步并非我们想象的那么容易，需要强大的算法来保障。在此列举两个因素为例进行说明：

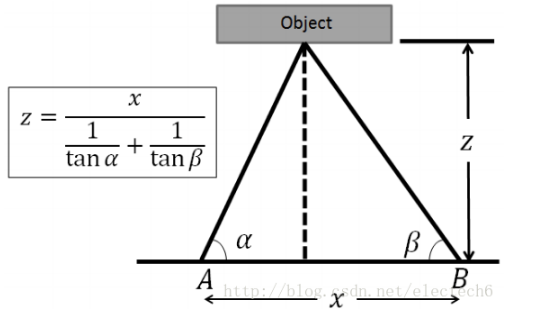
1、深度相机的分辨率目前还比较低，一般都是VGA（640 x 480）以下。而现在普通的RGB相机分辨率都已经到千万像素级以上了，是深度相机分辨率的几十倍甚至上百倍。因此需要将低分辨的深度图变成和RGB相机一致的高分辨率，这种“从无到有”的分辨率提升需要利用彩色图中物体的纹理、边界等内容信息，这个过程要想保持细节是比较困难的。

2、深度相机和彩色相机融合时还需要知道两个相机的畸变系数、光学中心、相对旋转/平移量等一系列参数，这就需要对两个相机进行相机标定工作。而深度相机的低分辨率对于相机标定工作也是一个较大的挑战。

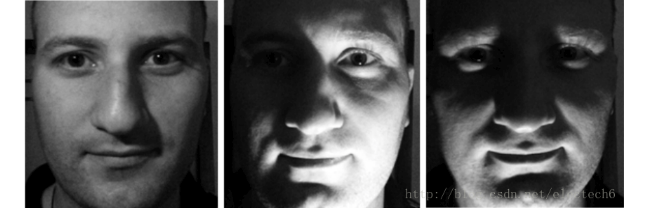
TOF Vs 双目视觉

读者可能会有疑问，现在双摄手机上的两个普通的彩色相机不就可以计算深度吗？和这个深度相机测距有何不同？

双目立体视觉测距的原理和人眼类似，通过计算空间中同一个物体在两个相机成像的视差就可以根据如下三角关系计算得到物体离相机的距离：

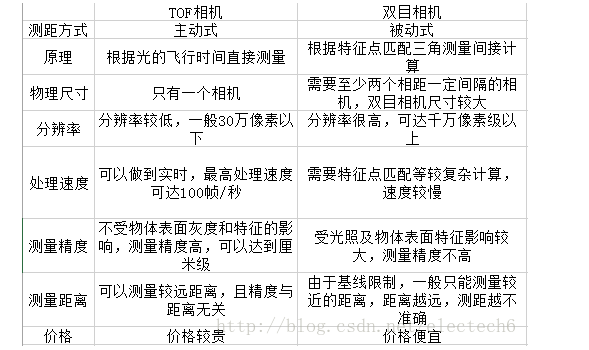


但是说起来容易，算法实现比较困难，双目立体视觉测距算法的核心就是寻找对应关系。可以理解为给定一个相机拍摄的图片中的任意一个像素点，如何在另外一个相机拍摄的图像中找到和它对应的像素点，这个过程需要特征提取、特征匹配等一系列复杂的算法。但是由于光照变化、光线明暗等外在因素的影响，拍摄的两张图片差别可能比较大，这会对特征匹配算法提出很大的挑战。如下图是在不同光照条件下拍摄的图片：



另外，如果拍摄的物体缺乏纹理和细节（比如拍摄一张白纸）的话，也很难进行特征匹配。这些都对算法的鲁棒性提出了很大的挑战。

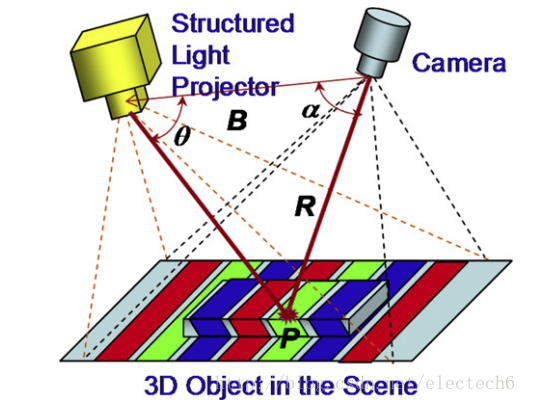
虽然TOF相机和双目立体视觉都能实现测距的功能，但是它们还是有很大不同，在此我做了了简单的表格如下：



结构光

简单的说一下结构光测距的方法吧！

结构光技术就是使用提前设计好的具有特殊结构的图案（比如离散光斑、条纹光、编码结构光等），然后将图案投影到三维空间物体表面上，使用另外一个相机观察在三维物理表面成像的畸变情况。如果结构光图案投影在该物体表面是一个平面，那么观察到的成像中结构光的图案就和投影的图案类似，没有变形，只是根据距离远近产生一定的尺度变化。但是，如果物体表面不是平面，那么观察到的结构光图案就会因为物体表面不同的几何形状而产生不同的扭曲变形，而且根据距离的不同而不同，根据已知的结构光图案及观察到的变形，就能根据算法计算被测物的三维形状及深度信息。

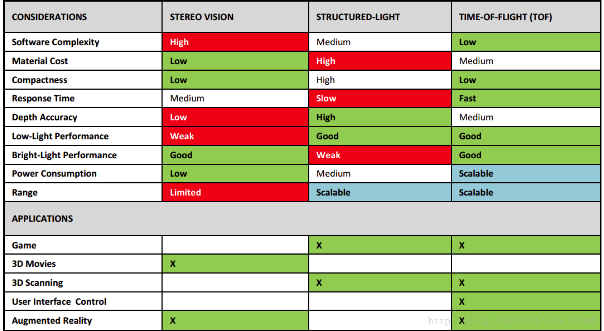


业界比较有名的就是以色列PrimeSense公司的Light Coding的方案，该方案最早被应用于Microsoft的明星产品Kinect上。目前该公司被苹果公司收购，可见苹果公司也将在深度相机领域有所动作。

结构光技术受环境光源影响较大，更适合室内的应用场景，而且帧率较低，所以更适合静态场景或者缓慢变化的场景。其优势就是能够获得较高分辨率的深度图像。

双目视觉 Vs 结构光 Vs TOF

下表是双目立体视觉、结构光、TOF三种可以测量深度（距离）的技术方案综合比较：



从上述的对比分析来看，TOF方案具有响应速度快，深度信息精度高，识别距离范围大，不易受环境光线干扰等优势。因此想要在移动端直接实现深度的测量，最有竞争力的就是TOF方案了。

典型手机

典型代表手机：联想Phab 2。

目前可以买到的具备直接深度测量的智能手机只有Google和联想合作的联想Phab 2，2016年11月推出，是全球首款支持Google Project Tango技术的手机，其深度相机采用TOF技术方案，由PMD公司提供。

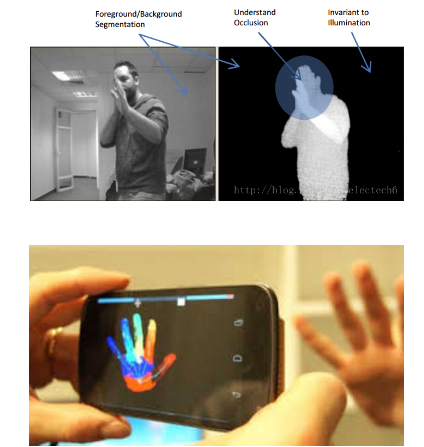
华硕也宣布将在2017年会推出带深度相机的手机Zenfone  AR，号称是全球首款同时支持Google Project Tango（AR）和Daydream（VR）的手机。

深度相机应用

深度相机的应用范围非常广泛：比如未来几年将会迅速商业化的手势识，以及活体人脸识别、空间测距、三维重建、AR(增强现实)等领域。

**1、手势识别。**

TOF深度相机可以将人脸、身体、手臂、手指从背景中分离，并且这种分割置信度较高，不受自然光变化的影响，同时能够实时性处理，所以这将在智能交互领域大有用武之地。预计最近几年会迅速进入消费级电子产品中。



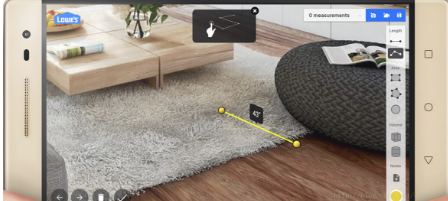
**2、真实的AR游戏体验。**

如下图是Phab 2的AR游戏展示。由于在二维图像中融合了实时的深度信息，所以AR游戏的体验比较真实。比如虚拟出来的一只猫，通过实时的空间深度感知，它可以“感受”到空间的相对位置关系，当它走到桌子边缘的时候，会很自然地跳到地面上，这在之前的AR游戏中是难以实现的。



**3、三维空间测量。**

由于能够实时获得深度信息，所以实现三维空间测量也是顺其自然的。比如在室内装修领域，可以方便的将各种虚拟的家具以真实的尺寸摆放到现实环境中，用户拿着手机就可以体验家居放在室内的360°真实效果，这无疑将是一个令人激动的应用场景。



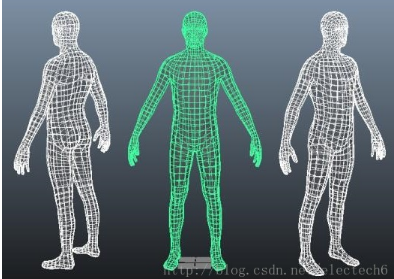
**4、三维扫描/重建。**

可以用于三维物体和k建模和机器人视觉导航和定位。比如你看到一座非常喜欢的雕塑，就可以利用手机上的彩色相机+深度相机对它扫描一周，结合相应的算法就可以生成该雕塑的三维模型数据，利用三维打印机就可以方便的打印出一个三维的雕塑复制品出来。



**5、更广泛的其他应用。**

融入了深度信息的三维影像可以用于活体人脸识别，避免传统二维人脸识别的安全隐患；可以更加方便进行人体三维测量，从而推动虚拟在线试衣行业的发展等。





随着深度测量技术的发展，必然还有出现更多有趣的应用场景。

继2016年双摄手机开始成为旗舰机的标配后后，今年领先的手机厂商或将联合上下游产业链将深度相机用于智能手机上，上述很多有趣的应用将会改变我们的生活，我们一起迎接这个技术日新月异的智能手机世界吧！