|  |  |  |
| --- | --- | --- |
| Lesson 7 Screen02  The Screen02 lesson builds on Screen01, by teaching how to draw lines and also a small feature on generating pseudo random numbers. It is assumed you have the code for the [Lesson 6: Screen01](http://www.cl.cam.ac.uk/projects/raspberrypi/tutorials/os/screen01.html) operating system as a basis.   |  | | --- | | **Contents**   * [1 Dots](http://www.cl.cam.ac.uk/projects/raspberrypi/tutorials/os/screen02.html#dots) * [2 Lines](http://www.cl.cam.ac.uk/projects/raspberrypi/tutorials/os/screen02.html#lines) * [3 Randomness](http://www.cl.cam.ac.uk/projects/raspberrypi/tutorials/os/screen02.html#randomness) * [4 Pi-casso](http://www.cl.cam.ac.uk/projects/raspberrypi/tutorials/os/screen02.html#picasso) |   1 Dots  Now that we've got the screen working, it is only natural to start waiting to create sensible images. It would be very nice indeed if we were able to actually draw something. One of the most basic components in all drawings is a line. If we were able to draw a line between any two points on the screen, we could start creating more complicated drawings just using combinations of these lines.  To allow complex drawing, some systems use a colouring function rather than just one colour to draw things. Each pixel calls the colouring function to determine what colour to draw there.  We will attempt to implement this in assembly code, but first we could really use some other functions to help. We need a function I will call SetPixel that changes the colour of a particular pixel, supplied as inputs in r0 and r1. It will be helpful for future if we write code that could draw to any memory, not just the screen, so first of all, we need some system to control where we are actually going to draw to. I think that the best way to do this would be to have a piece of memory which stores where we are going to draw to. What we should end up with is a stored address which normally points to the frame buffer structure from last time. We will use this at all times in our drawing method. That way, if we want to draw to a different image in another part of our operating system, we could make this value the address of a different structure, and use the exact same code. For simplicity we will use another piece of data to control the colour of our drawings.  Copy the following code to a new file called 'drawing.s'.  .section .data .align 1 foreColour: .hword 0xFFFF  .align 2 graphicsAddress: .int 0  .section .text .globl SetForeColour SetForeColour: cmp r0,#0x10000 movhs pc,lr ldr r1,=foreColour strh r0,[r1] mov pc,lr  .globl SetGraphicsAddress SetGraphicsAddress: ldr r1,=graphicsAddress str r0,[r1] mov pc,lr  This is just the pair of functions that I described above, along with their data. We will use them in 'main.s' before drawing anything to control where and what we are drawing.  Building generic methods like SetPixel which we can build other methods on top of is a useful idea. We have to make sure the method is fast though, since we will use it a lot.  Our next task is to implement a SetPixel method. This needs to take two parameters, the x and y co-ordinate of a pixel, and it should use the graphicsAddress and foreColour we have just defined to control exactly what and where it is drawing. If you think you can implement this immediately, do, if not I shall outline the steps to be taken, and then give an example implementation.   1. Load in the graphicsAddress. 2. Check that the x and y co-ordinates of the pixel are less than the width and height. 3. Compute the address of the pixel to write. (hint: frameBufferAddress + (x + y \* width) \* pixel size) 4. Load in the foreColour. 5. Store it at the address.   An implementation of the above follows.   1. .globl DrawPixel DrawPixel: px .req r0 py .req r1 addr .req r2 ldr addr,=graphicsAddress ldr addr,[addr] 2. height .req r3 ldr height,[addr,#4] sub height,#1 cmp py,height movhi pc,lr .unreq height  width .req r3 ldr width,[addr,#0] sub width,#1 cmp px,width movhi pc,lr   Remember that the width and height are stored at offsets of 0 and 4 into the frame buffer description respectively. You can refer back to 'frameBuffer.s' if necessary.   1. ldr addr,[addr,#32] add width,#1 mla px,py,width,px .unreq width .unreq py add addr, px,lsl #1 .unreq px   **mla dst,reg1,reg2,reg3** multiplies the values from **reg1** and **reg2**, adds the value from **reg3** and places the least significant 32 bits of the result in **dst**.  Admittedly, this code is specific to high colour frame buffers, as I use a bit shift directly to compute this address. You may wish to code a version of this function without the specific requirement to use high colour frame buffers, remembering to update the SetForeColour code. It may be significantly more complicated to implement.   1. fore .req r3 ldr fore,=foreColour ldrh fore,[fore]   As above, this is high colour specific.   1. strh fore,[addr] .unreq fore .unreq addr mov pc,lr   As above, this is high colour specific.  2 Lines  The trouble is, line drawing isn't quite as simple as you may expect. By now you must realise that when making operating system, we have to do almost everything ourselves, and line drawing is no exception. I suggest for a few minutes you have a think about how you would draw a line between any two points.  When programming normally, we tend to be lazy with things like division. Operating Systems must be incredibly efficient, and so we must focus on doing things as best as possible.  I expect the central idea of most strategies will involve computing the gradient of the line, and stepping along it. This sounds perfectly reasonable, but is actually a terrible idea. The problem with it is it involves division, which is something that we know can't easily be done in assembly, and also keeping track of decimal numbers, which is again difficult. There is, in fact, an algorithm called Bresenham's Algorithm, which is perfect for assembly code because it only involves addition, subtraction and bit shifts.  Algorithm derivation  Let's start off by defining a reasonably straightforward line drawing algorithm as follows:  /\* We wish to draw a line from (x0,y0) to (x1,y1), using only a function setPixel(x,y) which draws a dot in the pixel given by (x,y). \*/ if x1 > x0 then  set deltax to x1 - x0 set stepx to +1  otherwise  set deltax to x0 - x1 set stepx to -1  end if  if y1 > y0 then  set deltay to y1 - y0 set stepy to +1  otherwise  set deltay to y0 - y1 set stepy to -1  end if  if deltax > deltay then  set error to 0 until x0 = x1 + stepx  setPixel(x0, y0) set error to error + deltax ÷ deltay if error ≥ 0.5 then  set y0 to y0 + stepy set error to error - 1  end if set x0 to x0 + stepx  repeat  otherwise  /\* As above but swap x and y \*/  end if  This algorithm is a representation of the sort of thing you may have imagined. The variable error keeps track of how far away from the actual line we are. Every step we take along the x axis increases this error, and every time we move down the y axis, the error decreases by 1 unit again. The error is measured as a distance along the y axis.  While this algorithm works, it suffers a major problem in that we clearly have to use decimal numbers to store error, and also we have to do a division. An immediate optimisation would therefore be to change the units of error. There is no need to store it in any particular units, as long as we scale every use of it by the same amount. Therefore, we could rewrite the algorithm simply by multiplying all equations involving error by deltay, and simplifying the result. Just showing the main loop:  set error to 0 × deltay until x0 = x1 + stepx  setPixel(x0, y0) set error to error + deltax ÷ deltay × deltay if error ≥ 0.5 × deltay then  set y0 to y0 + stepy set error to error - 1 × deltay  end if set x0 to x0 + stepx  repeat  Which simplifies to:  set error to 0 until x0 = x1 + stepx  setPixel(x0, y0) set error to error + deltax if error × 2 ≥ deltay then  set y0 to y0 + stepy set error to error - deltay  end if set x0 to x0 + stepx  repeat  Suddenly we have a much better algorithm. We see now that we've eliminated the need for division altogether. Better still, the only multiplication is by 2, which we know is just a bit shift left by 1! This is now very close to Bresenham's Algorithm, but one further optimisation can be made. At the moment, we have an if statement which leads to two very similar blocks of code, one for lines with larger x differences, and one for lines with larger y differences. It is worth checking if the code could be converted to a single statement for both types of line.  The difficulty arises somewhat in that in the first case, error is to do with y, and in the second case error is to do with x. The solution is to track the error in both variables simultaneously, using negative error to represent an error in x, and positive error in y.  set error to deltax - deltay until x0 = x1 + stepx or y0 = y1 + stepy  setPixel(x0, y0) if error × 2 > -deltay then  set x0 to x0 + stepx set error to error - deltay  end if if error × 2 < deltax then  set y0 to y0 + stepy set error to error + deltax  end if  repeat  It may take some time to convince yourself that this actually works. At each step, we consider if it would be correct to move in x or y. We do this by checking if the magnitude of the error would be lower if we moved in the x or y co-ordinates, and then moving if so.  Bresenham's Line Algorithm was developed in 1962 by Jack Elton Bresenham, 24 at the time, whilst studying for a PhD.  Bresenham's Algorithm for drawing a line can be described by the following pseudo code. Pseudo code is just text which looks like computer instructions, but is actually intended for programmers to understand algorithms, rather than being machine readable.  /\* We wish to draw a line from (x0,y0) to (x1,y1), using only a function setPixel(x,y) which draws a dot in the pixel given by (x,y). \*/ if x1 > x0 then  set deltax to x1 - x0 set stepx to +1  otherwise  set deltax to x0 - x1 set stepx to -1  end if  set error to deltax - deltay until x0 = x1 + stepx or y0 = y1 + stepy  setPixel(x0, y0) if error × 2 ≥ -deltay then  set x0 to x0 + stepx set error to error - deltay  end if if error × 2 ≤ deltax then  set y0 to y0 + stepy set error to error + deltax  end if  repeat  Rather than numbered lists as I have used so far, this representation of an algorithm is far more common. See if you can implement this yourself. For reference, I have provided my implementation below.  .globl DrawLine DrawLine: push {r4,r5,r6,r7,r8,r9,r10,r11,r12,lr} x0 .req r9 x1 .req r10 y0 .req r11 y1 .req r12  mov x0,r0 mov x1,r2 mov y0,r1 mov y1,r3  dx .req r4 dyn .req r5 /\* Note that we only ever use -deltay, so I store its negative for speed. (hence dy**n**) \*/ sx .req r6 sy .req r7 err .req r8  cmp x0,x1 subgt dx,x0,x1 movgt sx,#-1 suble dx,x1,x0 movle sx,#1  cmp y0,y1 subgt dyn,y1,y0 movgt sy,#-1 suble dyn,y0,y1 movle sy,#1  add err,dx,dyn add x1,sx add y1,sy  pixelLoop$:  teq x0,x1 teqne y0,y1 popeq {r4,r5,r6,r7,r8,r9,r10,r11,r12,pc}  mov r0,x0 mov r1,y0 bl DrawPixel  cmp dyn, err,lsl #1 addle err,dyn addle x0,sx  cmp dx, err,lsl #1 addge err,dx addge y0,sy  b pixelLoop$  .unreq x0 .unreq x1 .unreq y0 .unreq y1 .unreq dx .unreq dyn .unreq sx .unreq sy .unreq err  3 Randomness  So, now we can draw lines. Although we could use this to draw pictures and whatnot (feel free to do so!), I thought I would take the opportunity to introduce the idea of computer randomness. What we will do is select a pair of random co-ordinates, and then draw a line from the last pair to that point in steadily incrementing colours. I do this purely because it looks quite pretty.  Hardware random number generators are occasionally used in security, where the predictability sequence of random numbers may affect the security of some encryption.  So, now it comes down to it, how do we be random? Unfortunately for us there isn't some device which generates random numbers (such devices are very rare). So somehow using only the operations we've learned so far we need to invent 'random numbers'. It shouldn't take you long to realise this is impossible. The operations always have well defined results, executing the same sequence of instructions with the same registers yields the same answer. What we instead do is deduce a sequence that is pseudo random. This means numbers that, to the outside observer, look random, but in fact were completely determined. So, we need a formula to generate random numbers. One might be tempted to just spam mathematical operators out for example: 4x2! / 64, but in actuality this generally produces low quality random numbers. In this case for example, if x were 0, the answer would be 0. Stupid though it sounds, we need a very careful choice of equation to produce high quality random numbers.  This sort of discussion often begs the question what do we mean by a random number? We generally mean statistical randomness: A sequence of numbers that has no obvious patterns or properties that could be used to generalise it.  The method I'm going to teach you is called the quadratic congruence generator. This is a good choice because it can be implemented in 5 instructions, and yet generates a seemingly random order of the numbers from 0 to 232-1.  The reason why the generator can create such long sequence with so few instructions is unfortunately a little beyond the scope of this course, but I encourage the interested to research it. It all centralises on the following quadratic formula, where xn is the nth random number generated.  xn+1 = axn2 + bxn + c mod 232  Subject to the following constraints:   1. a is even 2. b = a + 1 mod 4 3. c is odd   If you've not seen mod before, it means the remainder of a division by the number after it. For example b = a + 1 mod 4 means that b is the remainder of dividing a + 1 by 4, so if a were 12 say, b would be 1 as a + 1 is 13, and 13 divided by 4 is 3 remainder 1.  Copy the following code into a file called 'random.s'.  .globl Random Random: xnm .req r0 a .req r1  mov a,#0xef00 mul a,xnm mul a,xnm add a,xnm .unreq xnm add r0,a,#73  .unreq a mov pc,lr  This is an implementation of the random function, with an input of the last value generated in **r0**, and an output of the next number. In my case, I've used a = EF0016, b = 1, c = 73. This choice was arbitrary but meets the requirements above. Feel free to use any numbers you wish instead, as long as they obey the rules.  4 Pi-casso  OK, now we have all the functions we're going to need, let's try it out. Alter main to do the following, after getting the frame buffer info address:   1. Call SetGraphicsAddress with r0 containing the frame buffer info address. 2. Set four registers to 0. One will be the last random number, one will be the colour, one will be the last x co-ordinate and one will be the last y co-ordinate. 3. Call random to generate the next x-coordinate, using the last random number as the input. 4. Call random again to generate the next y-coordinate, using the x-coordinate you generated as an input. 5. Update the last random number to contain the y-coordinate. 6. Call SetForeColour with the colour, then increment the colour. If it goes above FFFF16, make sure it goes back to 0. 7. The x and y coordinates we have generated are between 0 and FFFFFFFF16. Convert them to a number between 0 and 102310 by using a logical shift right of 22. 8. Check the y coordinate is on the screen. Valid y coordinates are between 0 and 76710. If not, go back to 3. 9. Draw a line from the last x and y coordinates to the current x and y coordinates. 10. Update the last x and y coordinates to contain the current ones. 11. Go back to 3.   As always, a solution for this can be found on the downloads page.  Once you've finished, test it on the Raspberry Pi. You should see a very fast sequence of random lines being drawn on the screen, in steadily incrementing colours. This should never stop. If it doesn't work, please see our troubleshooting page.  When you have it working, congratulations! We've now learned about meaningful graphics, and also about random numbers. I encourage you to play with line drawing, as it can be used to render almost anything you want. You may also want to explore more complicated shapes. Most can be made out of lines, but is this necessarily the best strategy? If you like the line program, try experimenting with the SetPixel function. What happens if instead of just setting the value of the pixel, you increase it by a small amount? What other patterns can you make? In the next lesson, [Lesson 8: Screen 03](http://www.cl.cam.ac.uk/projects/raspberrypi/tutorials/os/screen03.html), we will look at the invaluable skill of drawing text. | 第7课 屏幕02  这一课是建立在第6课的基础上的，而且主要教授如何绘制线以及生成伪随机数的小窍门。这里假设你具有了第6课的代码。  目录   1. 点 2. 线 3. 随机 4. 毕加索   1. 点  既然我们已经可以让屏幕工作了，何不现在就是产生一个看得见的图片呢？的确，如果我们可以绘制些东东，那将是多么美妙。在所有的绘制操作中最基本的就是绘制线。如果我们可以在屏幕上的任意两点间绘制一条线段，那么我们就可以用这些线的组合绘制出更加复杂的任何图形。  为了绘制复杂图形，一些系统使用一个颜色函数来绘制，而不是只用一种颜色。每个像素的绘制过程中都要调用颜色函数来决定该点的颜色。  我们将使用汇编代码来实现这个函数，但是首先一些其他的函数来帮助我们。我们需要一个名叫SetPixel的函数来改变一个特定像素的颜色数值，寄存器r0和r1的数值作为该函数的输入参数。如果将来我要写一段在任何内存位置绘制图形，而不是仅仅在屏幕上绘制的话，这将会很有用。因此，首先，我需要一个系统去控制我要绘制的地方。我认为最好的办法就是利用一片内存来存储我们要绘制地方的信息。该函数将以存储了一个最后指向帧缓存结构体的指针的地址来终结。我们将会在我们绘图函数中时刻用到该函数。如果我们想要在我们的操作系统中另外一个地方来绘制一张不同的图的话，我们就可以使用那个不同的结构体的地址数值，而代码还是这段代码。简单起见，我们还使用另外一片数据去控制我们要绘制的图形的颜色。  把下面的代码拷贝并复制到一个新的文件里，该文件名为“drawing.s”。  .section .data .align 1 foreColour: .hword 0xFFFF  .align 2 graphicsAddress: .int 0  .section .text .globl SetForeColour SetForeColour: cmp r0,#0x10000 movhs pc,lr ldr r1,=foreColour strh r0,[r1] mov pc,lr  .globl SetGraphicsAddress SetGraphicsAddress: ldr r1,=graphicsAddress str r0,[r1] mov pc,lr  这仅仅是我上面描述的函数的带数据部分。我们会在文件“main.s”中使用该函数。在绘制任何图形之前，我们使用上面的代码来控制在哪里绘制，绘制什么。  创建一个类似于SetPixel的泛型函数，以便我们以后可以在其上面创建别的函数是个很有用的主意。我们必须确保该函数的速度要快，因为我们将会大量地使用到它。  我们的下一个任务就是去实现这个SetPixel函数。这需要两个参数，一个是像素的x和y坐标值，另外一个是使用graphicsAddress和foreColour来明确指明我们要绘制什么以及在哪里绘制。如果你自信满满，那么你就独立地实现它吧。如果还没有那么自信，我将描绘出要执行的大致步骤，然后给出一个样例。  1.加载graphicsAddress。  2.检测像素的x和y坐标是否小于width和height。  3.计算输入的像素的地址。（提示：公式为：frameBufferAddress + (x + y \* width) \* pixel size）。  4.加载foreColour。  5.把其地址存储起来。  下面是其一个样例实现。   1. .globl DrawPixel DrawPixel: px .req r0 py .req r1 addr .req r2 ldr addr,=graphicsAddress ldr addr,[addr] 2. height .req r3 ldr height,[addr,#4] sub height,#1 cmp py,height movhi pc,lr .unreq height  width .req r3 ldr width,[addr,#0] sub width,#1 cmp px,width movhi pc,lr   记住：严格来讲，width和height的信息存储在帧缓存描述体的偏移地址为0和4的地方。你可以回头参考一下文件“framebuffer.s”。   1. ldr addr,[addr,#32] add width,#1 mla px,py,width,px .unreq width .unreq py add addr, px,lsl #1 .unreq px   指令mla dst, reg1, reg2, reg3将会把寄存器reg1和reg2的数值进行乘积，而后和寄存器reg3中的数值相加，然后把结果的最低的32个位存储到dst中去。  无可否认，这段代码特定于颜色帧缓存的高地址，因为我使用了一个直接位移动来计算地址。你可能想要去编写一个不是特定于要求去使用颜色帧缓存高地址的代码，没问题，记得更新函数SetForeColour的代码就可以了。这可能是最复杂的实现了。  4. fore .req r3 ldr fore,=foreColour ldrh fore,[fore]  就像上面提到的，这段代码是特定于颜色的高地址。  5. strh fore,[addr] .unreq fore .unreq addr mov pc,lr  就像上面提到的，这段代码是特定于颜色的高地址。  2. 线  你可能对画线这个操作的困难程度估计不够。到目前为止，你必须明确一点的是，当我们要做一个操作系统的时候，我们几乎不得不亲自做所有的事情，画线操作也不例外。我的建议是，在继续学习之前，请认真思考一下，在任意两点之间绘制一条线段应该具体怎么做。  当我们编写普通的程序时，我们总是很懒得深入事物的内部，而仅仅把他们作为可以信赖的组件。所以，操作系统必须极其高效。因此，我们在写操作系统时，必须把眼光聚焦于做任何功能都要尽可能做到最好。  我期望大多数策略的中心观点要包括计算线段的倾角，然后一步步逼近它。这听起来很自然而然，但是它却是一个相当糟糕的主意。问题所在就是它的操作过程中包括除法，而这一点在汇编语言中无法很容易地实现。而且还要持续跟踪小数，这更增加了难度。事实上，存在一个叫做布兰森汉姆的算法。这个算法非常适合于汇编语言，因为它只包含加法，减法和位移。  算法的演进  布兰森汉姆线产生算法是由杰克.埃尔顿.布兰森汉姆于1962年开发的，那年他24岁，正在进行博士生学习期间。  布兰森汉姆线产生算法可以被描述位下面的伪代码。伪代码仅仅是一段看起来像代码的文本而已，它的主要目的是帮助程序员理解算法，而不是让机器执行的。  /\* We wish to draw a line from (x0,y0) to (x1,y1), using only a function setPixel(x,y) which draws a dot in the pixel given by (x,y). \*/ if x1 > x0 then  set deltax to x1 - x0 set stepx to +1  otherwise  set deltax to x0 - x1 set stepx to -1  end if  set error to deltax - deltay until x0 = x1 + stepx or y0 = y1 + stepy  setPixel(x0, y0) if error × 2 ≥ -deltay then  set x0 to x0 + stepx set error to error - deltay  end if if error × 2 ≤ deltax then  set y0 to y0 + stepy set error to error + deltax  end if  repeat  相比于 |