

HOG和SIFT圖像特徵提取簡述

計算機視覺life 昨天

點擊上方“

快速獲得最新乾貨

本文轉載自咪付，文章僅用於學術分享。

前言

如今，計算機視覺的應用在生活中已隨處可見，如人臉識別考勤門禁、全態識別過閘乘地鐵等。之所以能大量應用，是因為能解決問題，其本質原理是在數學上找到了一種從具體圖像轉換到特定數學空間的方法。這裡的特定數學空間稱為特徵空間，該轉換方法即為特徵提取方法。

常見的特徵提取的方法有兩大類：一類是通過傳統機器學習算法提取，另一類是通過深度神經網絡進行提取。在大數據處理方面，深度神經網絡具有更高的精準度、更廣的適用領域；而在數據量相對較小的領域，傳統機器學習算法速度精準度更具有優勢，這是因為傳統機器學習算法推理嚴格、計算過程可控。

本文將關注機器學習兩種經典的特徵提取算法：HOG和SIFT方法。

1、圖像識別與特徵

1.1 圖像識別

圖像識別是計算機視覺中的一個重要方向。簡單地說，圖像識別就是給一張圖片讓計算機認出其中的內容，包括目標的位置、語義、類別等信息。圖像有靜態和動態之分，這裡主要討論的是靜態圖像，理解了靜態圖像的識別原理就容易擴展到動態圖像。如下圖，計算機通過圖像識別給出圖像中有哪些目標（水果）、各目標在圖像中的位置等。



1.2 圖像特徵

計算機看圖像實際上是通過圖像的一系列像素最終辨別圖像有什麼，是什麼。當然，計算機不是直接從圖像的原始像素值就能辨識出圖像的內容，圖像的原始像素需經過運算後，抽像出圖像特徵才能“看”清楚圖像。圖像特徵的提取是非常關鍵的一環。

我們將通過分析兩種典型的可解析性擬合算法HOG、SIFT，來直觀了解圖像特徵提取過程。為什麼選擇HOG、SIFT，因為兩者俱有高度的相似性，都使用了梯度計算、直方圖統計。

兩者大體步驟對比如下：

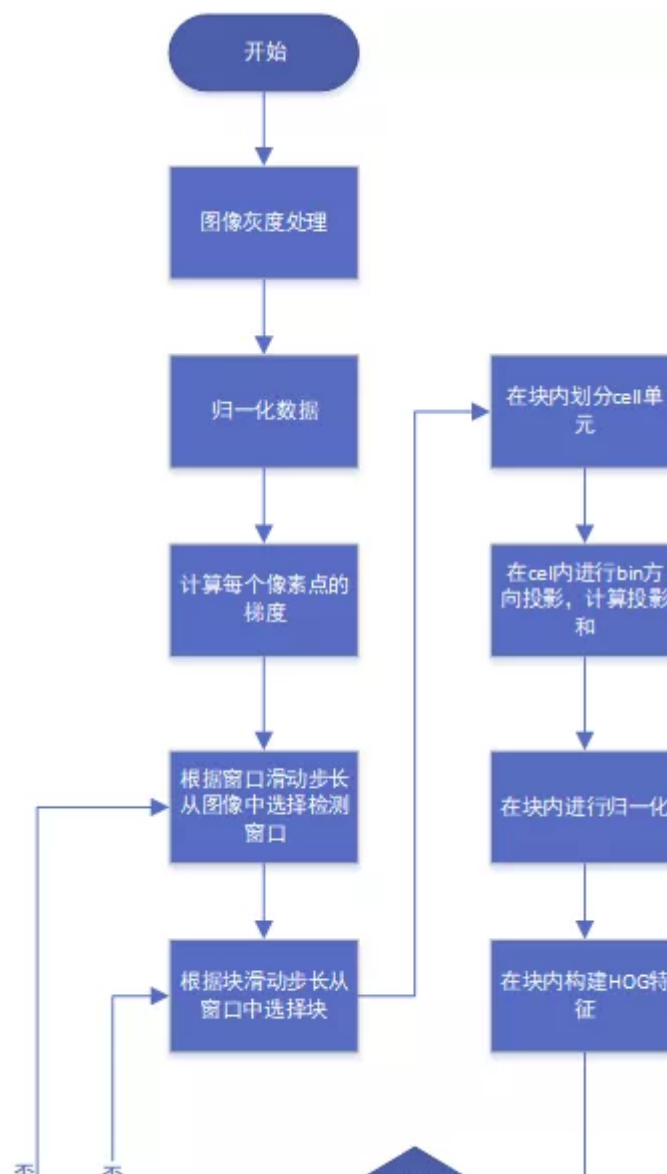
步驟	HOG	SIFT
預處理1	圖像灰度化	圖像縮放（多個不同分辨率的圖像）
預處理2	圖像歸一化	高斯模糊、DoG特徵圖計算
描述子計算	分塊計算	計算極值點、分塊計算

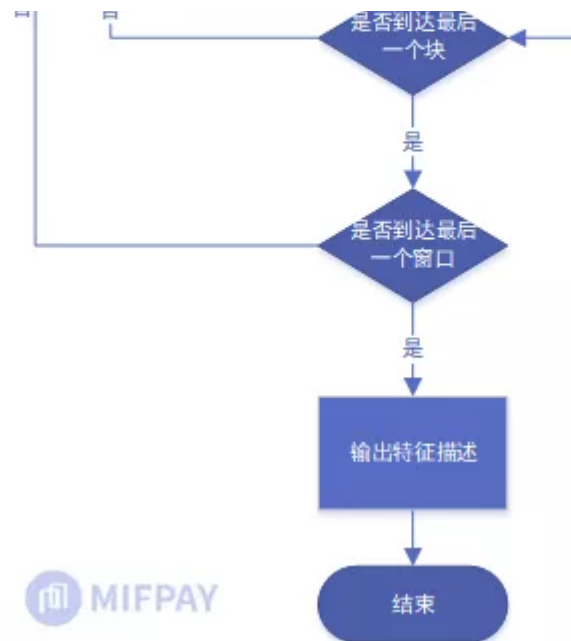
兩者特點對比如下：

對比項	HOG	SIFT
特徵特點	單一	多尺度
應用領域	目標檢測 目標跟踪	圖像匹配 三維建模
優點	圖像幾何和光學形變都保持良好的不變性、剛性物體特徵提取效果好	尺度不變性、旋轉不變性、亮度不變性、噪點不敏感、特徵維度小、抗遮擋
缺點	特徵維度大、描述子生成過程冗長、無法處理遮擋、噪點相當敏感	對邊緣不明顯

2、HOG特征

2.1 特征提取流程





2.2 圖像灰度化

彩色圖像一般有3個通道，分別是R（紅色）、G（綠色）、B（藍色），像素取值範圍為0-255。灰度圖像就只有1個通道。將彩色圖像變成灰度圖像，通用的灰度化公式為： $\text{gray} = 0.3 \times R + 0.59 \times G + 0.11 \times B$ 。

假設一幅RGB圖像為：240×240像素，將它所有像素值組成的向量維度是：（240×240 × 3）維。灰度化後，其所有像素值組成的向量維度則是：240×240 維，維度為RGB圖像的三分之一，這就是灰度化的意義之一。

2.3 圖像歸一化

圖像歸一化是為了減少光照等的影響，降低圖像局部的陰影，避免在圖像的紋理強度中，局部表層曝光較大的情況。歸一化的類別有：gamma空間歸一化和顏色空間歸一化。

標準化的Gamma壓縮公式為：

$$I(x, y) = I(x, y)^{Gamma}$$

根據不同的效果要求，gamma可取不同的值，常見取值有：1/2、對數等。

2.4 梯度幅值和梯度方向

計算圖像橫坐標和縱坐標方向的梯度，並根據橫坐標和縱坐標的梯度，計算梯度大小和梯度方向。

圖像中某個像素點(x,y)的水平方向梯度和垂直方向梯度的計算式分別為：

$$\begin{aligned} G_x(x, y) &= H(x+1, y) - H(x-1, y) \\ G_y(x, y) &= H(x, y+1) - H(x, y-1) \end{aligned}$$

通過該像素點水平方向梯度和垂直方向梯度計算其梯度大小和方向分別如下：

$$\begin{aligned} G(x, y) &= \sqrt{G_x^2(x, y) + G_y^2(x, y)} \\ \theta(x, y) &= \tan^{-1} \frac{G_y(x, y)}{G_x(x, y)} \end{aligned}$$

在算法中，通常利用梯度算子（例如：[-1,0,1]）對原圖像做卷積運算，得到水平方向（x）的梯度值，再利用轉置的梯度算子對原圖像做卷積運算，得到豎直方向（y）的梯度值。最後通過上述公式計算該像素點的梯度大小和方向。

典型的梯度算子如下圖所示：

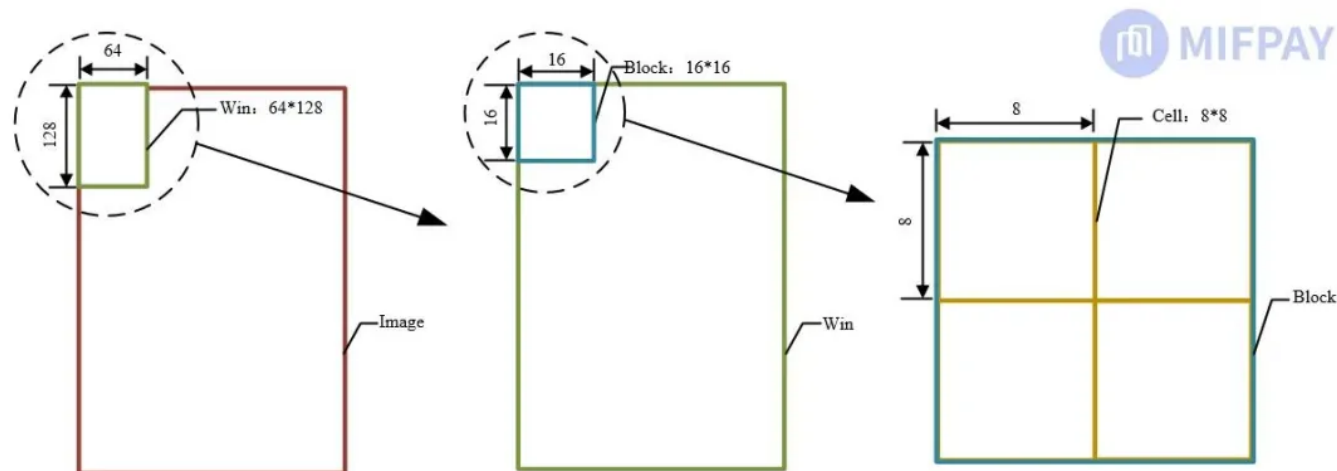


梯度橫坐標方向算子與縱坐標方向算子

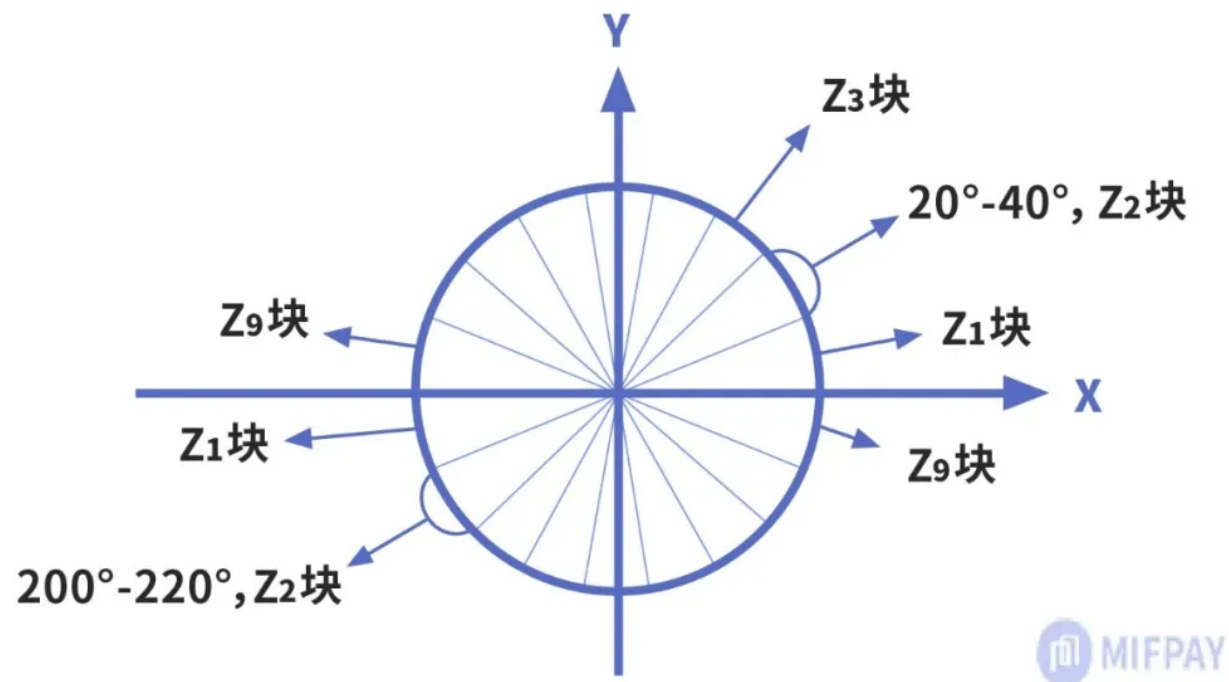
梯度算子可以是多種多樣的，例如還可以是 $[-1, 1]$ 、 $[1, -8, 0, 8, -1]$ 等等。

2.5 窗口 (win) 、塊 (block) 、胞元 (cell) 和箱 (bin)

我們首先將灰度圖像像素點劃分成一個或多個窗口 (win)，例如劃分win的大小為：64*128。一個窗口再分為多個塊 (block)，例如劃分block的大小為：16*16。一個塊再分為多個細胞單元 (cell)，例如劃分cell的大小為：8*8。

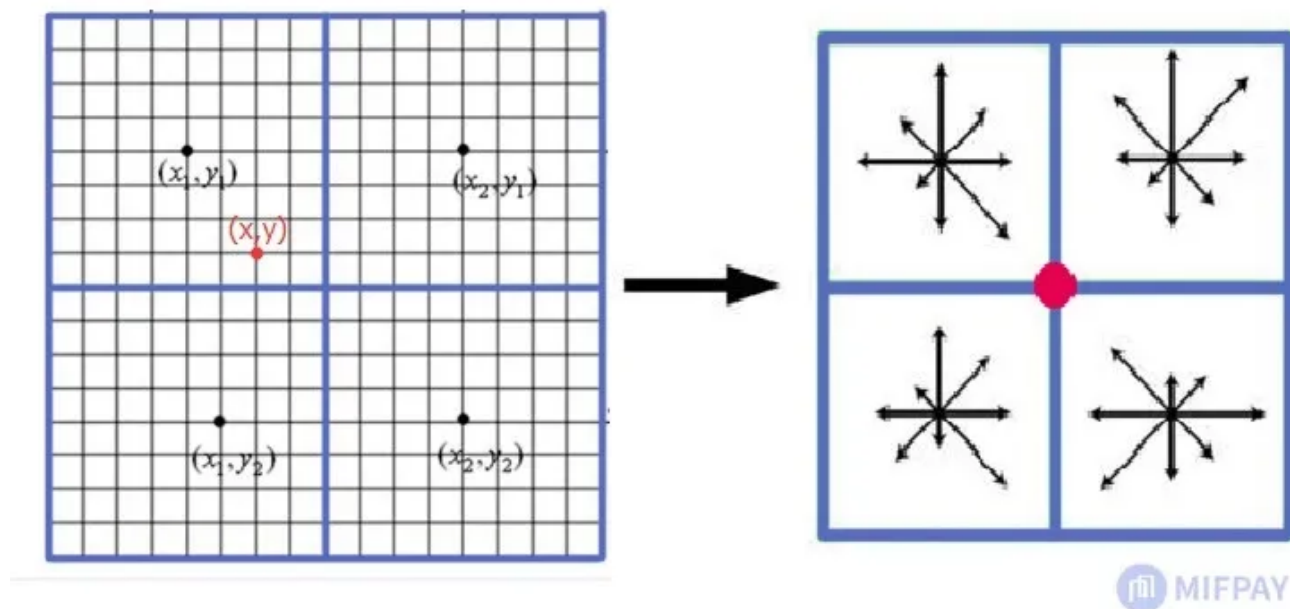


將梯度方向按一定角度分開，通常在0-180度（無向）或0-360度（有向）範圍內。例如採用無向的梯度和9個直方圖通道，則方向的範圍劃分為 $180/9=20$ 度，方向即劃分為9個箱（bin），如下圖示，有 $z_1 \sim z_9$ 共9個bin。



z1 ~ z9的9個bin的值代表該梯度方向上累加的梯度幅值。假如該cell區域中某個像素梯度方向為15°，梯度幅值為2，則在cell梯度向量第1個bin(Z1塊)中值加2（不考慮插值）。對每個cell區域內的所有像素按其梯度方向循環累加，得到該cell區域的梯度向量值。

如下圖所示：



一個窗口 (win) 的總HOG特徵數為：block總數× block中特徵數，即：

$$\{(window_width / step - 1) * (window_height / step - 1)\} * \{num_of_bin * (block_width/cell_width) * (block_height/cell_height)\}$$

例如在前面的例子中：

win: 64x128; block: 16x16; cell: 8x8; 步長: 8; bin個數: 9。

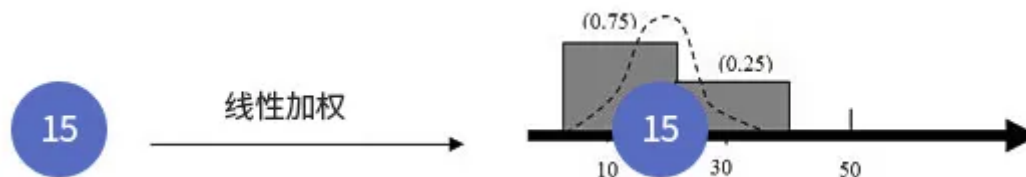
則block總數為： $(64/8 - 1) * (128/8 - 1) = 105$,

那麼一個win的總特徵數為： $105 * (9 * 2 * 2) = 3780$ 。

2.6 插值

當計算出來的特徵向量没能精準的匹配到某個bin上時，也就是說該點的特徵向量在多個bin上都能投影，為了準確地該特徵點在每個bin上的投影，根據插值的思想得將梯度幅值乘以對應的權重後再累加到相應的bin上。

權值的計算方式，根據圖像處理中圖像放大算法插值的思想，線性插值如下圖所示：



以前面的例子為例，我們將直方圖分成了9個直方圖通道，方向的範圍劃分為 $180/9=20$ 度，對應為9個bin，每個bin為20度。對梯度方向為 15° 像素點進行處理，通過計算可以得出，該點跟Z1 ($0\sim 20^\circ$)、Z2 ($20^\circ\sim 40^\circ$) 塊bin最為接近。那麼該 15° 像素點在Z1、Z2塊bin上的權值分別為（以Z1、Z2的中心點計算權值）： $(30 - 15) / 20 = 0.75$ 、 $(15 - 10) / 20 = 0.25$ 。

同理將原圖像的每個像素點跟根據插值的方式計算出權值，然後將像素點的幅值映射到整個bin中，即可完成整個bin直方圖統計。

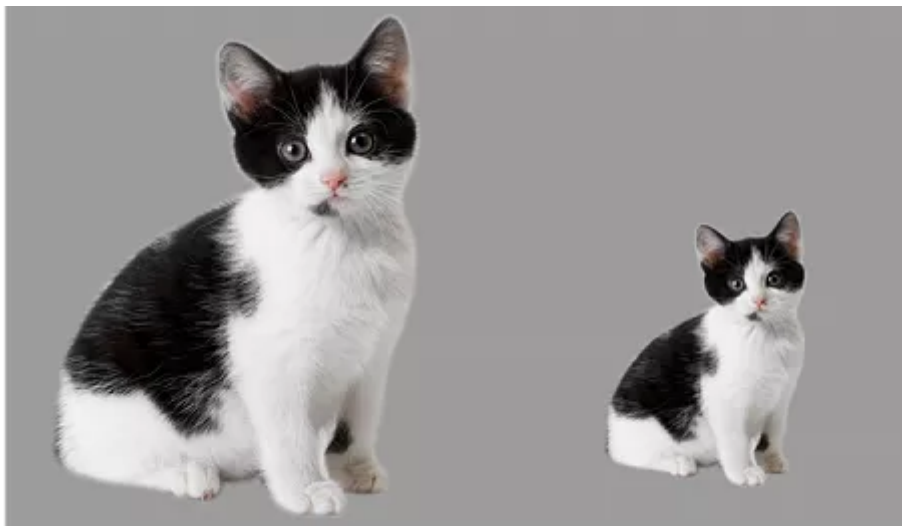
2.7 塊內歸一化操作

由於局部光照的變化以及前景-背景對比度的變化，使得梯度強度的變化範圍非常大。這就需要對梯度強度做歸一化。歸一化能夠進一步地對光照、陰影和邊緣進行壓縮，使特徵向量空間對光照，陰影和邊緣變化具有魯棒性。具體做法是把各個細胞單元組合成大的、空間上連通的區域（blocks）。這樣，一個block內所有cell的特徵向量串聯起來便得到該block的HOG特徵。這些區間是互有重疊的，這就意味著：每一個單元格的特徵會以不同的結果多次出現在最後的特徵向量中。歸一化之後的塊描述符（向量）就稱之為HOG描述符。

3、SIFT算法

SIFT (Scale-invariant feature transform) 特徵提取算法，即尺度不變性特徵變換，是Lowe在1999年提出的，並於2004年對其進行了完善總結。傳統CV算法中，SIFT的應用一直居於榜首，足見其分量。

SIFT算法確立的特徵點穩定性很好，不會因為尺度變化、旋轉變化、光照的變化、噪音的干擾而受影響，這也是SIFT特徵提取算法非常突出的優點。例如下圖展示了尺度變化的圖片，在左邊放大的圖片中採用SIFT算法確立的特徵點與右邊縮小的圖片確立的特徵點能很好實現匹配。



SIFT算法的實現步驟總的來說可分為兩步：特徵點檢出和特徵點描述（類似HOG的過程）。

3.1 特徵點檢出

SIFT算法的第一步是找到足夠多的特徵點。主要是用了DoG，就是把圖像做不同程度的高斯模糊，平滑的區域一般變化不大，而紋理複雜的比如邊緣、點、角之類區域一般變化很大，這樣變化很大的點就是特徵點。當然為了找到足夠的點，還需要把圖像放大、縮小幾倍來重複這個步驟找特徵點。

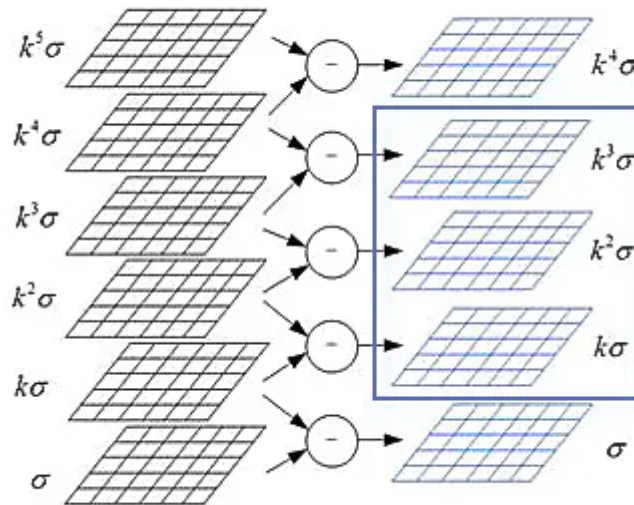
3.11 構建多尺度DoG空間

首先，構建6張圖片，分別是原圖經不同方差高斯模糊後的圖像，方差如下：

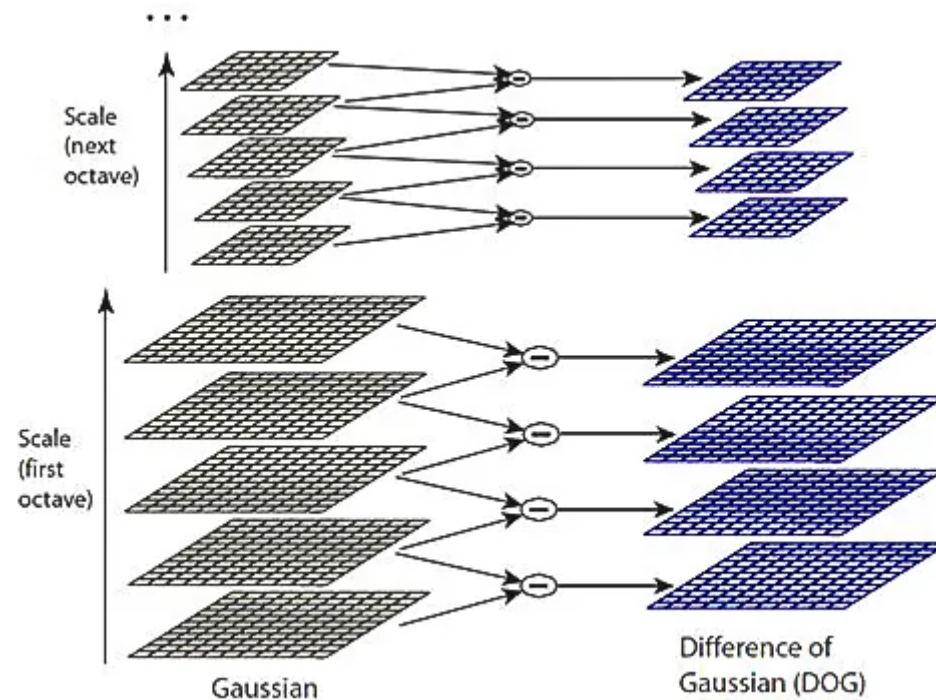
$$\sigma, k\sigma, k^2\sigma, k^3\sigma, k^4\sigma, k^5\sigma$$

經上述方差濾波後，從左到右，圖像越來越模糊。

其次，構建DoG圖片。DoG圖就是由上述經不同方差高斯模糊後的圖像相鄰兩張之差得出。如下圖所示，6張經高斯模糊的圖片相鄰兩張之差得到5張DoG圖。

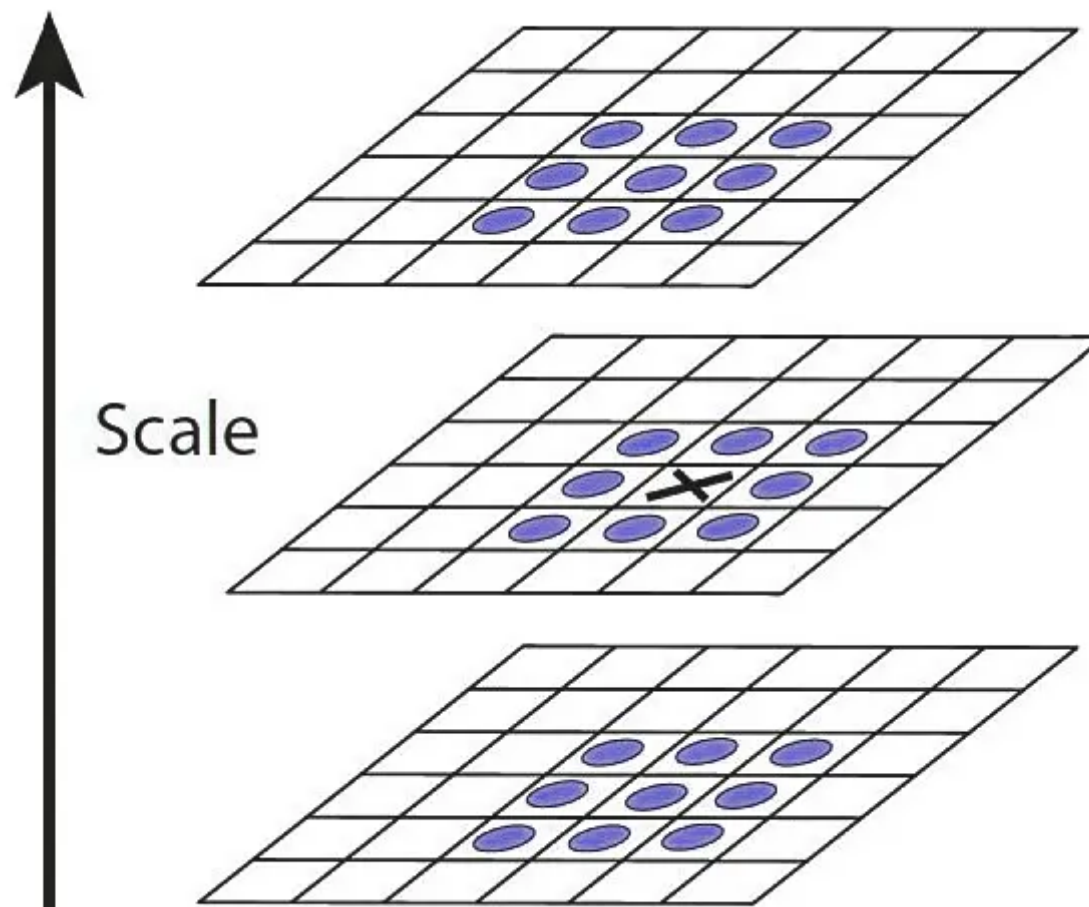


SIFT算法解決尺度不變性，就是把上述過程拓展到多分辨率上，下圖就是由兩種分辨率構成的DoG空間。相同分辨率的照片組成一個尺度空間，每個尺度空間由5張遞增高斯模糊處理的照片構成。以此類推，拓展到更多分辨率上時構成了多尺度DoG空間。

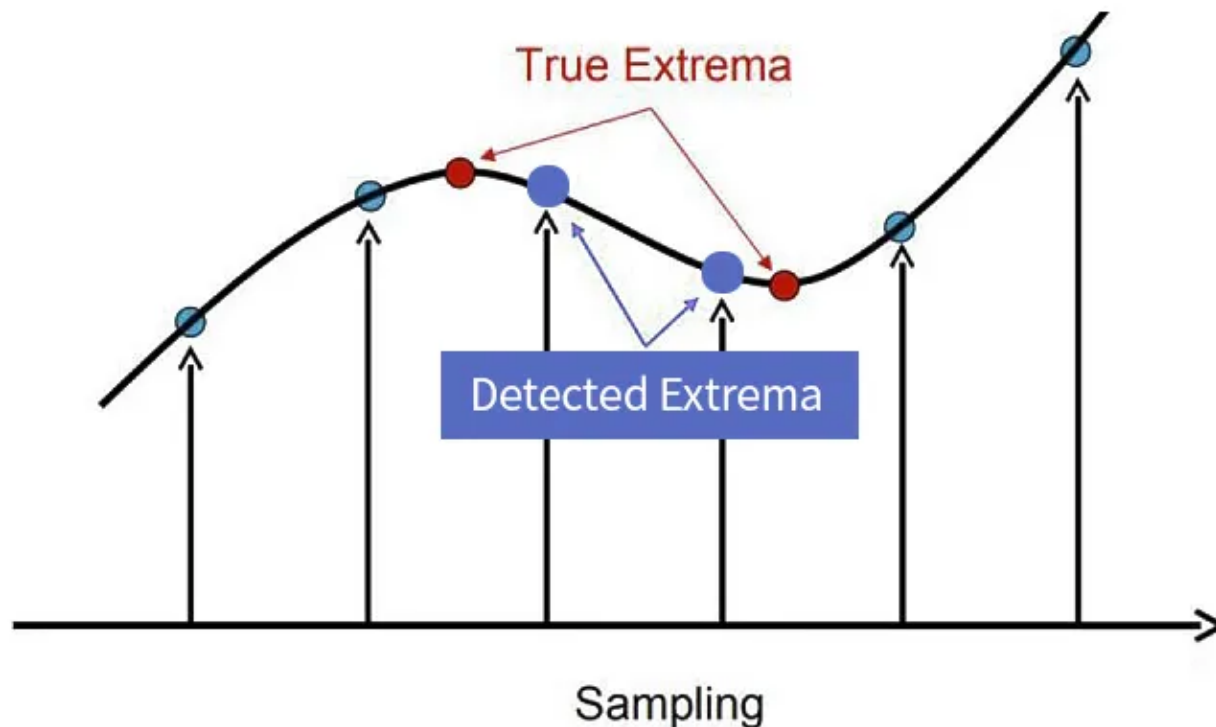


3.12 尋找極值點（特徵點）

接下來，在多尺度DoG空間中尋找到極致點。怎麼找？比較法。一幅DoG圖像，讓目標點跟它周圍的8鄰域的8個點比較，並且跟它相鄰尺度的DoG圖像做三維的空間比較，因此一個目標點總共會同周邊26個點比較，如下圖。如果目標點經過如此比較後，確實是這26個點中的極大或極小值，就認為該點是圖像在該尺寸下的極值點。



按照上述比較法，找到多尺度DoG空間的所有極值點。但這些極值點都是離散的，在離散的空間中，局部極值點可能並不是真正意義上的極值點。因此，採用插值法進一步求得真正的極值點。如下圖所示，利用已知的離散空間點插值得到連續空間極值點，通過對尺度空間DoG函數進行曲線擬合，進一步對函數方程求偏導，得到精確的極值點。



此外，還需要刪除邊緣效應的點。DoG的值會受到邊緣的影響，那些邊緣上的點，雖然不是斑點，但是它的DoG響應也很強。一旦特徵點落在圖像的邊緣上，這些點就是不穩定的點。一方面圖像邊緣上的點是很難定位的，具有定位歧義性；另一方面這樣的點很容易受到噪聲的干擾而變得不穩定，所以要把這部分點刪除。我們利用橫跨邊緣的地方，在沿邊緣方向與垂直邊緣方向表現出極大與極小的主曲率這一特性。所以通過計算特徵點處主曲率的比值即可以區分其是否在邊緣上。

3.2 獲取特徵點描述

檢出足夠多的特徵點後，我們就可以開始計算這些特徵點的特徵得到對應特徵點的描述。這一步和HOG算法類似，即以檢出的特徵點為中心選16x16的區域作為local patch，這個區域又可以均分為4x4個子區域。對於每一個關鍵點，都擁有位置、尺度以及方向三個信

息。為每個關鍵點建立一個描述符，用一組向量將這個關鍵點描述出來，使其不隨各種變化而改變，比如光照變化、視角變化等等。這個描述子不但包括關鍵點，也包含關鍵點周圍對其有貢獻的像素點，並且描述符應該有較高的獨特性，以便於提高特徵點正確匹配的概率。

3.21 選取特徵點的方向

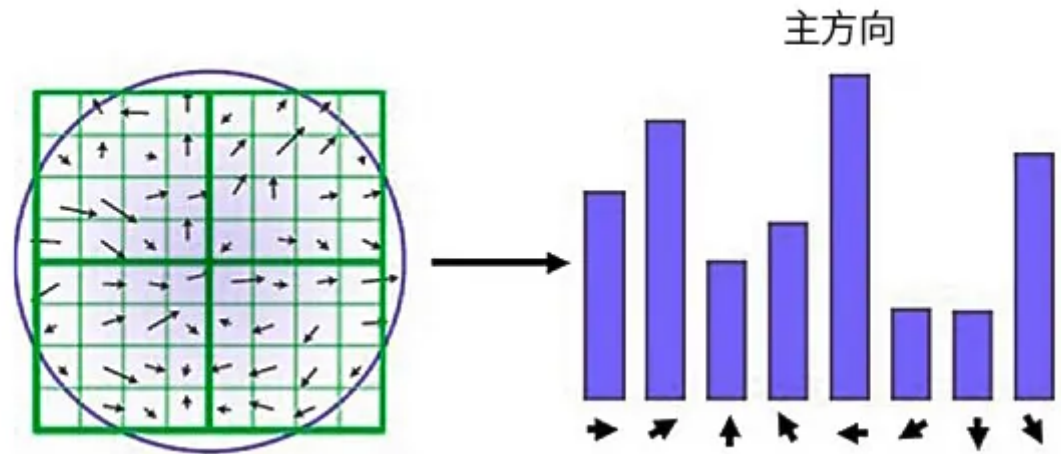
採集特徵點所在高斯金字塔圖像 3σ 鄰域窗口內像素的梯度和方向分佈特徵。梯度的幅值和方向如下：

$$m(x, y) = \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2}$$

$$\theta(x, y) = \tan^{-1}((L(x, y+1) - L(x, y-1)) / (L(x+1, y) - L(x-1, y)))$$

在完成特徵點的梯度計算後，使用直方圖統計鄰域內像素的梯度和方向。子區域中各個像素的梯度都可以分到8個bin（類似HOG算法）裡面。

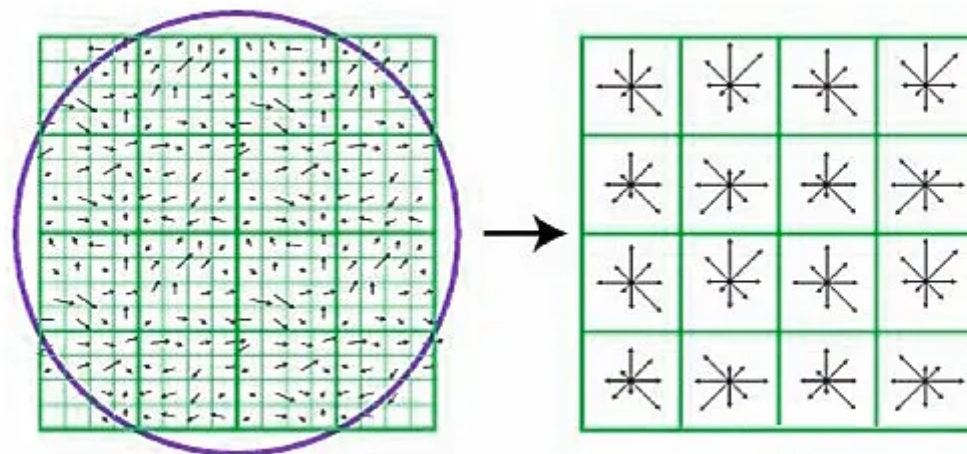
梯度直方圖將0~360度的方向範圍分為36個柱，其中每柱10度。如圖所示，直方圖的峰值則代表了該特徵點處鄰域梯度的主方向，即作為該特徵點的方向，其他的達到最大值80%的方向可作為輔助方向。



3.22 選取特徵點的方向

SIFT描述子是特徵點鄰域高斯圖像梯度統計結果的一種表示。通過對特徵點周圍圖像區域分塊，計算塊內梯度直方圖，生成具有獨特性的向量，這個向量是該區域圖像信息的一種抽象，具有唯一性。

特徵描述子與特徵點所在的尺度有關，因此，對梯度的求取應在特徵點對應的高斯圖像上進行。將關鍵點附近的鄰域劃分為 $d*d$ (Lowe建議 $d=4$)個子區域，每個子區域做為一個種子點，每個種子點有8個方向。



每一個小格都代表了特徵點鄰域所在的尺度空間的一個像素，箭頭方向代表了像素梯度方向，箭頭長度代表該像素的幅值。然後在 4×4 的窗口內計算8個方向的梯度方向直方圖。繪製每個梯度方向的累加可形成一個種子點。這樣就可以對每個特徵點形成一個 $4 \times 4 \times 8 = 128$ 維的描述子。

4、特徵的使用

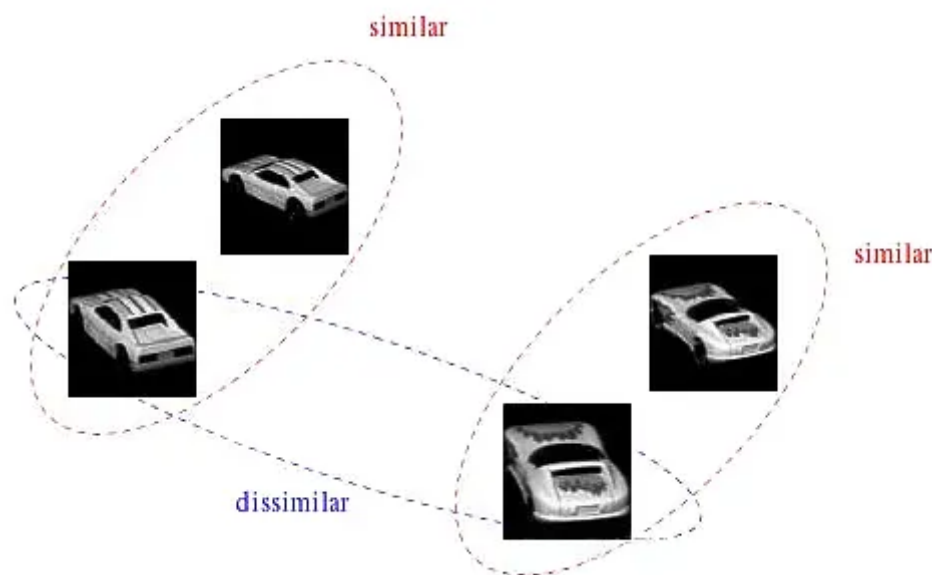
特徵的提取是為了解決問題，計算機視覺的問題分為兩個經典類別：分類、回歸。

4.1 分類

分類是人臉識別中最常用的一種手段，即特徵數據匹配，即在數據庫中找到最相似的特徵數據，即認為該最相似的特徵數據對應的圖像與我們要識別的圖像最接近，從而判斷出人臉圖像的ID。

HOG特徵通常結合SVM分類器在行人識別中應用效果非常好。將檢測窗口中所有重疊的塊進行HOG特徵採集，並將它們結合成最終的特徵向量供分類使用。分類算法常採用SVM分類器SVMLight。

如下圖所示：在特徵檢索匹配時，我們認為同一物體的不同圖像是相近的（同一輛車的不同時刻、角度的圖像），而同類的物體之間是相近的（都是車的圖像）。理想的圖像匹配算法應該認為兩幅同一物體的圖像之間相似度很高，而兩幅不同物體的圖像之間相似度很低。



4.2 回歸

回歸問題在計算機視覺中的主要應用是感知出各個目標的位置、語義等信息（環境感知）。是目標檢測常用的算法，起到的作用是預測boundingbox的信息。深度學習領域的yolo系列、retinanet等one-state目標檢測，以及rcnn系列等two-state目標檢測，整體的框架基本上都是特徵提取+回歸。

當然回歸這裡所說的特徵提取就不僅是前面介紹的HOG、SIFT特徵，更多的是深度神經網絡輸出的特徵。

5、總結

特徵提取是計算機視覺算法中非常重要的一環，根據解決問題的不同、特徵提取的方法也是多種多樣。不管是人工特徵、傳統機器學習、深度學習，本質都是擬合，只是擬合路徑的方式不同，由人工設計到可解釋擬合到只關心輸入輸出的自動擬合。

深度神經網絡的擬合上限比傳統學習的上限高，HOG、SIFT可以視為擬合路徑的特定路徑，這個特定路徑方法有邏輯性。深度網絡擬合路徑沒有限制，通過訓練的方式找到合適擬合路徑。

也是因為深度神經網絡擬合上限很大以及傳統機器學習擬合路徑可控的原因，在有限數據樣本上，傳統的機器學習的表現能力有很多方面是強於深度學習的。

獨家重磅課程！

- 1、VIO課程：
- 2、圖像三維重建課程（第2期）：
- 3、基於LiDAR的多傳感器融合SLAM 系列教程：LOAM、LeGO-LOAM、LIO-SAM
- 4、系統全面的相機標定課程：
- 5、視覺SLAM必備基礎（第2期）：
- 6、深度學習三維重建課程：
- 7、激光定位+建圖課程：手把手教你Cartographer從入門到精通！



扫描学习SLAM、三维重建

全國最棒的SLAM、三維視覺學習社區↓



技術交流微信群

歡迎加入公眾號讀者群一起和同行交流，目前有 請按照格式備註，否則不予通過 添加成功後會根據研究方向邀請進入相關微信群。請勿在群內發送
投稿、合作



掃描關注視頻號，看最新技術落地及開源方案視頻秀 ↓



扫一扫二维码，关注我的视频号

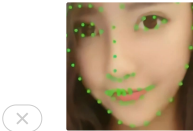
— 版權聲明 —

本公眾號原創內容版權屬計算機視覺life所有；從公開渠道收集、整理及授權轉載的非原創文字、圖片和音視頻資料，版權屬原作者。如果侵權，請聯繫我們，會及時刪除。

喜歡此內容的人還喜歡

雙一流博士整理的計算機視覺學習路線（深度學習+傳統圖像處理）

小白學視覺



2021年深度學習哪些方向比較新穎，處於上升期或者朝陽階段，比較有研究潛力？

計算機視覺聯盟

图像处理计算机视觉深度学习

向比较新颖，处于上升期？

感觉现在比较成熟饱和了，再研究趋势不是很了解。求大佬解答~

✚ 美图回答👍 对问题 318