HISTOGRAMS OF ORIENTATION GRADIENTS

graphics • vision • visualisation

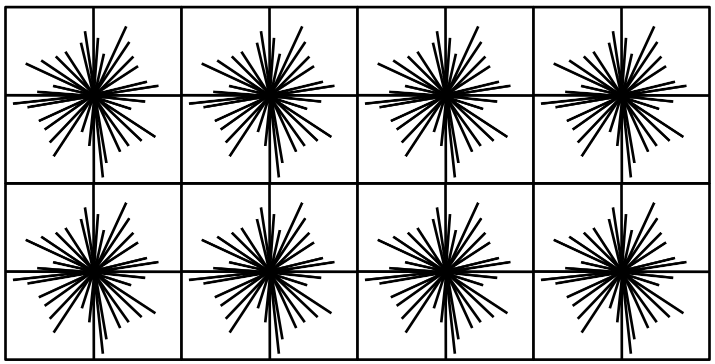
School of Computer Science & Statistics
Trinity College Dublin
Dublin 2
Ireland
www.scss.tcd.ie

OVERVIEW

- Advanced Feature Extraction
  - Viola-Jones (& Ada boost)✓
  - Optical Flow (& Motion analysis)✓
  - Histograms of Orientation Gradients - this lecture
  - Histograms of Flows - this lecture
  - Scale Invariant Feature Transforms - subsequent lecture

Histograms of Orientation Gradients

- Objective: object recognition
- Basic idea
  - Local shape information often well described by the distribution of intensity gradients or edge directions even without precise information about the location of the edges themselves.



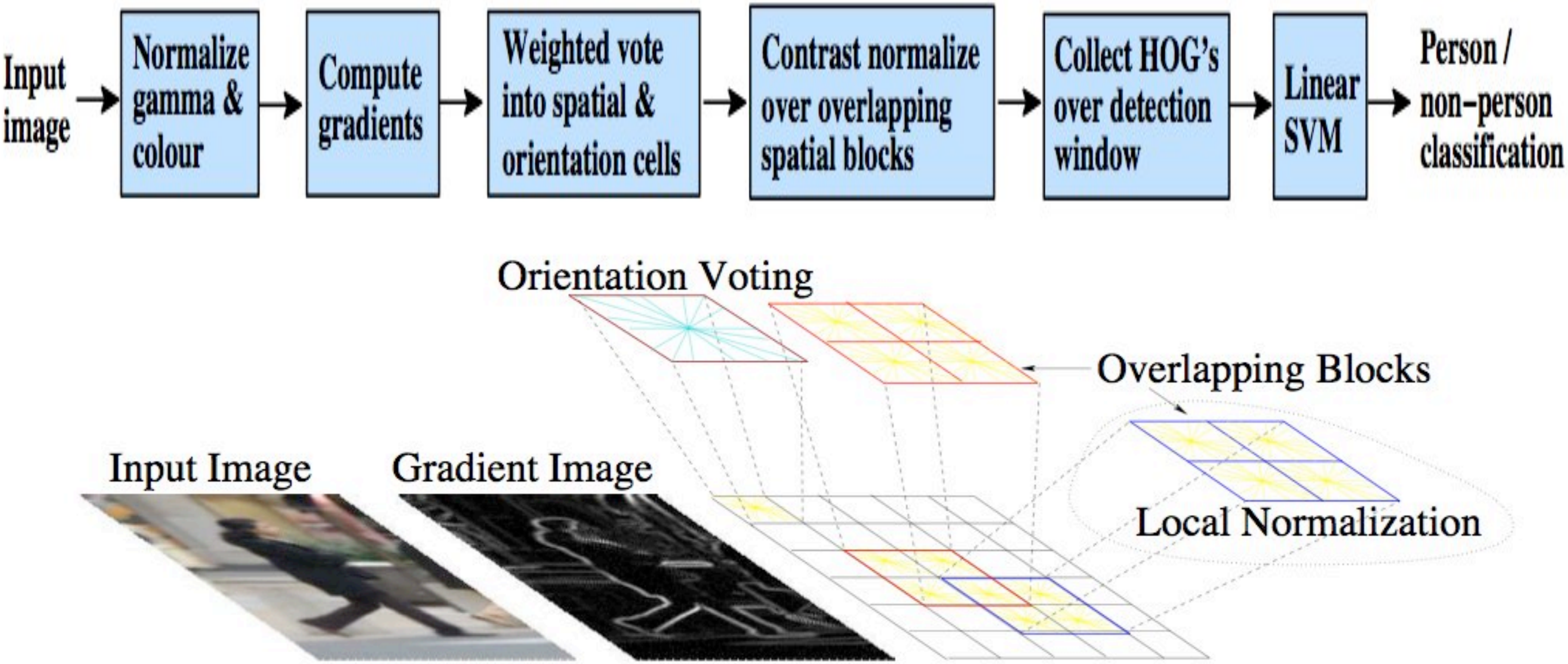
Algorithm Overview

- Divide image into small sub-images: “cells”
  - Cells can be rectangular (R-HOG) or circular (C-HOG)
- Accumulate a histogram of edge orientations within that cell
- The combined histogram entries are used as the feature vector describing the object
- To provide better illumination invariance (lighting, shadows, etc.) normalize the cells across larger regions incorporating multiple cells: “blocks”

Why HOG?

- Capture edge or gradient structure that is very characteristic of local shape
- Relatively invariant to local geometric and photometric transformations
  - Within cell rotations and translations do not affect the HOG values
  - Illumination invariance achieved through normalization
- The spatial and orientation sampling densities can be tuned for different applications
  - For human detection (Dalal and Triggs) coarse spatial sampling and fine orientation sampling works best
  - For hand gesture recognition (Fernandez-Llorca and Lacey) finer spatial sampling and orientation sampling is required

A worked example: Dalal and Triggs CVPR'05



Colour Normalisation

- RGB and LAB colour spaces equivalent
- Gamma correction (gain) no major affect on results
- Greyscale only small negative effect (-1.5%) on results

Computing the Gradient

- Several gradient detectors tried
  - [1,-1], [1,0,-1], [1,-8,0,8,-1], Sobel
  - Unfiltered and Pre-filtered with Gaussian smoothing
- Simplest [1,0,-1] proved best
- Gaussian smoothing affected results negatively
- For colour images
  - Compute each channel separately
  - Choose the largest value as the gradient for that pixel

Orientation Binning

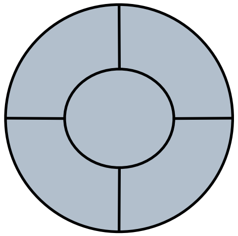
- Each pixel votes for an orientation according to the closest bin in the range
  - 0 to 180 (ignore negative edge directions)
  - 0 to 360
- Bilinear smoothing to reduce aliasing effects
- The “vote” is weighted by the gradient magnitude
  - Magnitude, Magnitude^2, edge presence absence, etc

Normalization

- Gradient is affected by illumination changes > normalization needed
  - Normalization takes the maximum range of a signal and stretches it to take up the maximum possible range
    - e.g.a simple normalization scheme: if the range of pixel values is 50-180 out of a total of 0-255 then min=50 and max =180, normalization > (old_pix - min)*((max-min)/255)= new_pix
  - Other normalization schemes can be used
- Group cells into larger blocks
- Normalize within the blocks
  - This ensures that low contrast regions are stretched
- Overlapping blocks
  - This ensures consistency across the whole image without the loss of local variations

Block Geometries

- R-HOG
  - Rectangular arrangement of cells
  - E.g. 6x6 cells
- C-Hog
  - Circular arrangement of cells



Classification

- Detector Window
  - 64X128 Dalal and Triggs
  - 128X128 hand washing
- Support Vector Machine(SVM) Classifier
  - We will discuss these later in the course.
- Feature vector made up of the multiple HOGs within the detector window
  - Shape information encoded by HOGs
  - Spatial location implicitly encoded by relative positions of HOG's within the detector window.

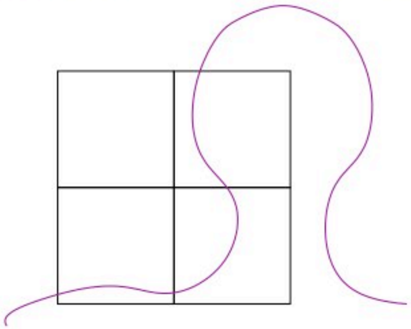
Dalal and Triggs results


input image


weighted pos wts


weighted neg wts


avg. grad


outside in block

- The most important cues are head, shoulder, leg silhouettes
- Vertical gradients inside the person count as negative
- Overlapping blocks those just outside the contour are the most important

Example video

Extracting Movement and Shape: histograms of flows

Dalal, Triggs and Schmid - ECCV 06

Hand Washing Analysis - Lacey, Llorca, Vilariño and Zhou

graphics • vision • visualisation

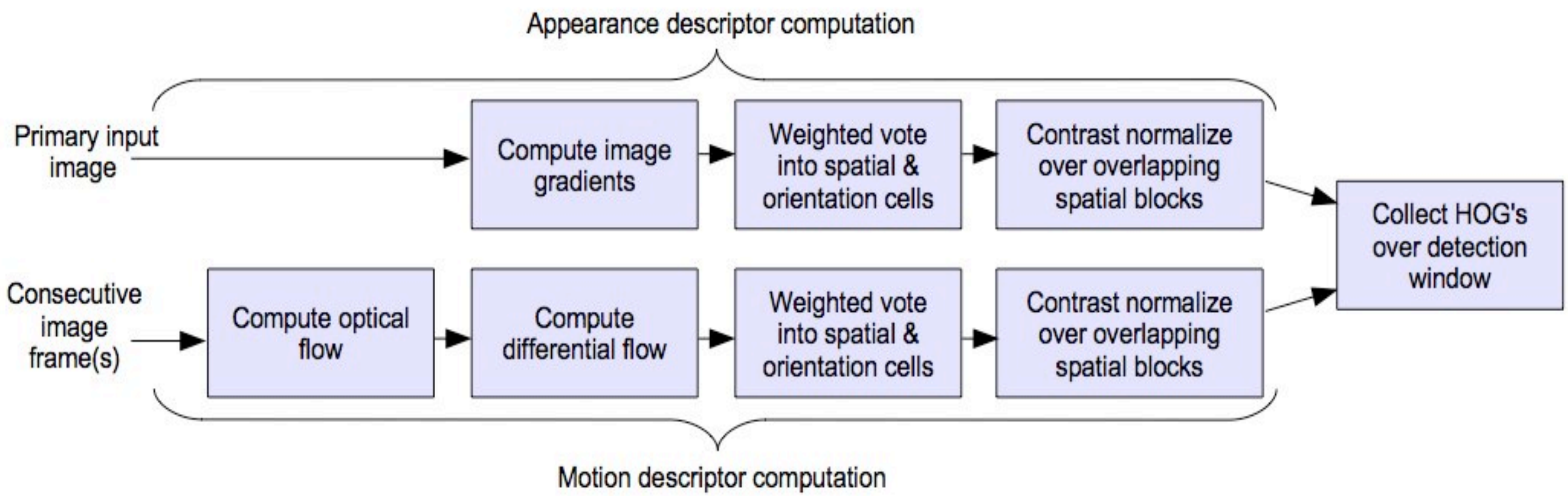
School of Computer Science & Statistics
Trinity College Dublin
Dublin 2
Ireland
www.scss.tcd.ie

Background

- Motion is a good feature but difficult to use
- Motion on its own is a weak classifier
- Combining motion with appearance or shape makes a strong classifier
- Earlier approaches
  - Viola, Jones and Snow 2003
    - harr wavelets and block matching using AdaBoost
  - Haritaoglu et.al 2000
    - Optical flow based activity recognition

Add motion to HOG

- Calculate optical flows
- Compute differential optical flow
- Combine in spatial and orientation cells as before



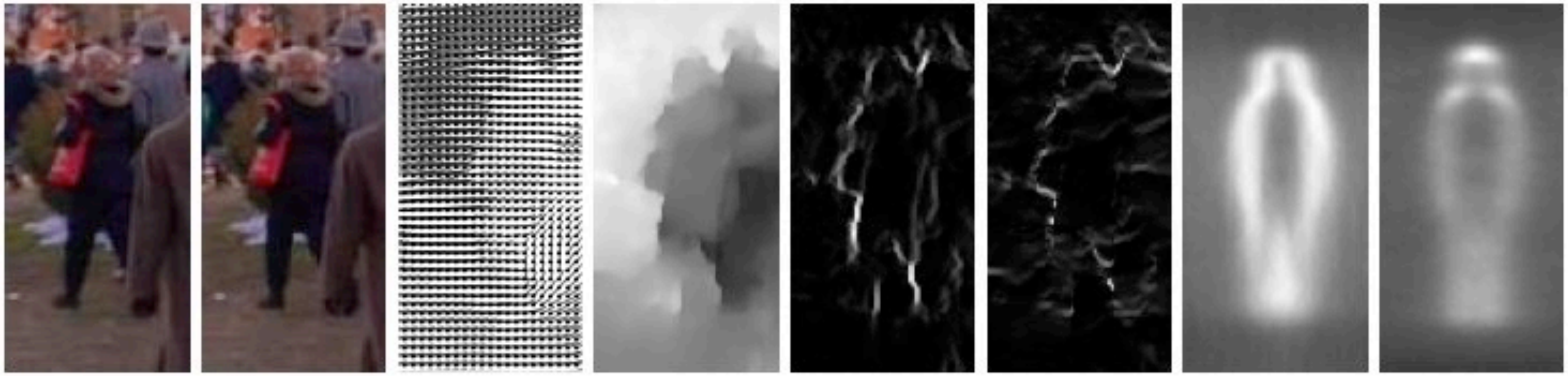
What type of motions?

- Camera Motions
  - Optical flow based on camera motions (pan, tilt and roll) is smooth over the image
  - Computing the difference of flows (first difference of optical flows) removes this.
- Object Motions
  - What remains are effects due to motion parallax due to changes in depth - object silhouette
  - Internal flows will give limb body relative motions

Motion Boundary Histogram (MBH)

- Optical flow gives two “images” the horizontal and vertical components of optical flow: I^x, I^y
- Treating each separately find the first derivative using a simple filter , [-1,0,1], with no smoothing
- Compute the combined histogram using winner takes all voting for each pixel
  - Similar to how colour was handled in the HOG case
- Or separate histograms can be built for each “image”

Example motions captured



(a) (b) (c) (d) (e) (f) (g) (h)

Fig. 3. Illustration of the MBH descriptor. (a,b) Reference images at time t and $t+1$. (c,d) Computed optical flow, and flow magnitude showing motion boundaries. (e,f) Gradient magnitude of flow field $\mathcal{I}^x, \mathcal{I}^y$ for image pair (a,b). (g,h) Average MBH descriptor over all training images for flow field $\mathcal{I}^x, \mathcal{I}^y$.

Internal Motion Histograms

- MBH captures shape - too similar to HOG?
- To capture “internal” relative motions use IMH family of descriptors
  - IMHdiff > different scales and angles
    - five: [-1,0,0,0,1] or seven: [-1,0,0,0,0,0,1] and diagonals also
  - IMHcd > split block into 3X3 cells, motion differences are calculated on a pixel by pixel basis subtracting the central pixel value from the corresponding pixel in each of the 8 outer cells
  - IMHmd > 3X3 normalised as full block
  - IMHwd > like IMHcd but uses Harr wavelets rather than differences to calculate values
  - IMHsd > viola-jones type of spatial differencing between cells in the blocks

Evaluations for people finding

Detector Set-up for detecting people

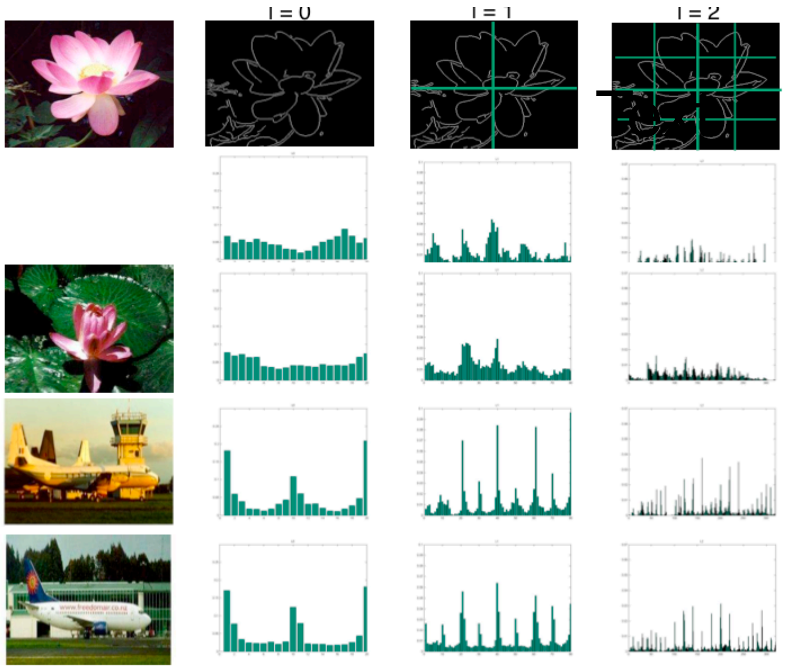
- Cell size 8x8
- 81 histogram bins per cell
- 2x2 blocks used for normalisation
- Horn and Schunk method used for optical flow calculation
- Static HOG + IMHcd performed best
- Complimentary Information in two detectors

Extensions

- Pyramidal HOG (PHOG)
  - *Representing shape with a spatial pyramid kernel* - CIVR-07
  - Bosch (Girona), Zisserman(Oxford), Munoz(Girona)
- 3D HOG
  - *3D Extended Histogram of oriented Gradients (3DHOG) for the classification of road users in urban scenes* - BMVC-09
  - Buch, Orwell, Velastin (Kingson)

PHOG

- Classify Images based on objects they contain
- Objective:
  - An image feature that changes depending on object type

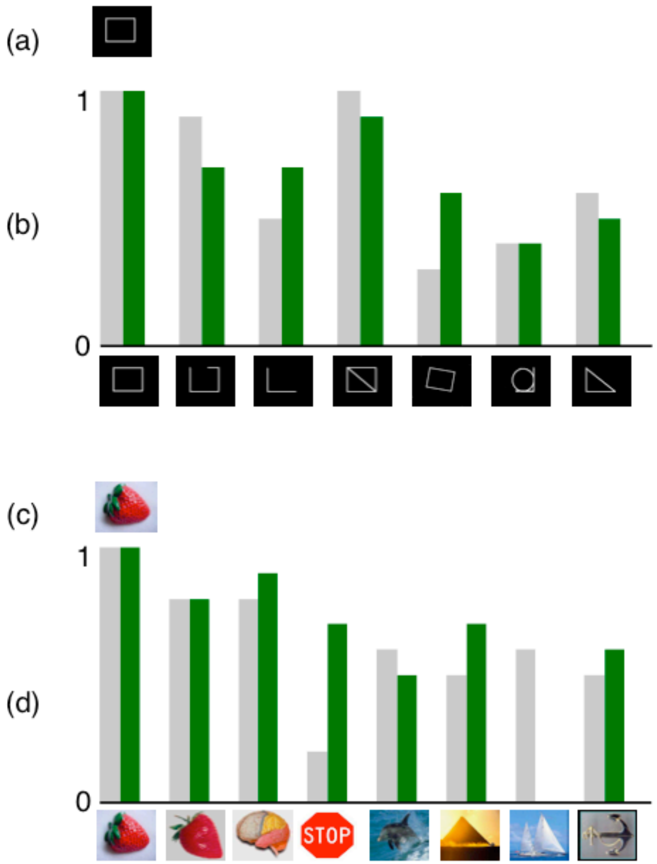


Different levels of grid

Weighted combination

Different object classes

PHOG performance



- Comparison with Chamfer matching
  - [Using χ^2 statistic]
- Why bother?
  - Tolerance of small rotations
  - Compact feature vector for use with machine learning
  - Less need for spatial matching
- More on machine learning later...

3DHOG

- 3D objects change appearance significantly depending on orientation
- Associate 2D HOG with surface patches of 3D model during training

Conclusions

- Shape and appearance important cues
- Motion *may* be an important additional cue
  - Motion vectors need to be dense
    - Block matching is too coarse
- Motion information needs to be complimentary to shape
  - People in moving images
    - “Internal” motion differencing
- Further Reading (links)
    - Histograms of Oriented Gradients for Human Detection Navneet Dalal and Bill Triggs
    - Human Detection Using Oriented Histograms of Flow and Appearance Navneet Dalal, Bill Triggs, and Cordelia Schmid