COMPUTER VISION REVISION

PHONG = AMBIENT + DIFFUSE + SPEC

VERTICAL EXAMPLES:
$\begin{bmatrix}1&1&1\\0&0&0\\-1&-1&-1\end{bmatrix}$ $\begin{bmatrix}1&2&1\\0&0&0\\-1&-2&-1\end{bmatrix}$ $\begin{bmatrix}1&1&1\\1&-2&1\\-1&-1&-1\end{bmatrix}$ $\begin{bmatrix}3&3&3\\3&0&3\\-5&-5&-5\end{bmatrix}$ LAPLACE = 2ND DERIVATIVE $\begin{bmatrix}0&1&0\\1&-4&1\\0&1&0\end{bmatrix}$ $\begin{bmatrix}1&1&1\\1&-8&1\\1&1&1\end{bmatrix}$

EDGE DETECTION = PREWITT, SOBEL, ROBINSON, KIRSCH

GRADIENT MAGNITUDE $\sqrt{g_x^2+g_y^2}$ GRADIENT DIR = ARCTAN$\left(\frac{g_y}{g_x}\right)$ (GRADIENT PERPENDICULAR TO EDGE)

HOUGH TRANSFORM = DESCRIBE A LINE IN M,C SPACE $y=mx+c$ OR IN R,θ SPACE (DISTANCE/ANGLE FROM ORIGIN)

WHERE CURVES OF DIFFERENT POINTS MEET, THEY ARE COLINEAR

HOUGH TRANSFORM FOR ELLIPSES CENTRE X_c, Y_c, $h = a/b$, θ (ANGLE, ECCENTRICITY, SIZE (a, b))

SEGMENT INTENSITIES BY HISTOGRAM

QUADTREE 1) SPLIT NON-UNIFORM SQUARES INTO 4 SUB-SQUARES UNTIL UNIFORM
2) MERGE NEIGHBORING BLOCKS THAT ARE SIMILAR

TEXTURE MAPPING - MATCH ENTROPY ($\Sigma x \log x$), CONTRAST, HOMOGENEITY

TRY HSV SPACE AS HUE DOESN'T VARY MUCH IN HIGHLIGHTS/SHADOWS

DESCRIBING A SHAPE BY MOMENTS $m_{pq} = \sum_i \sum_j i^p j^q f(i,j)$

M_{00} = AREA $X_c = \frac{M_{10}}{M_{00}}$ $Y_c = \frac{M_{01}}{M_{00}}$ SECOND MOMENTS SAY HOW SPREAD IT IS

FOURIER DESCRIPTOR DISTANCE AS FUNCTION OF ANGLE

FOURIER TRANSFORM CONVERTS BETWEEN FREQ & SPATIAL DOMAINS

HUMANS PERCEIVE PHASE > AMP

MULT IN FREQ = CONVOLUTION IN SPACE $F(u)G(u) \Leftrightarrow f(x)*g(x)$

CONVOLUTION ⇒ $h(i,j) = f(x)*g(x) = \sum_M \sum_N f(m,n)\, g(i-m, j-n)$

CORRELATION CAN MATCH SOURCE IMAGE TO TARGET

BAG OF FEATURES = EXTRACT FEATURES
MAKE DICTIONARY OF PATTERNS (WORDS)
CHARACTERIZE IMAGES BY FREQUENCY OF PATTERNS IN DICTIONARY

HARRIS CORNER DETECTOR - ROTATION INVARIANT

$E(u,v) = \sum_x \sum_y (uI_x + vI_y)^2 = [u, v] \begin{bmatrix}\Sigma I_x^2 & \Sigma I_x I_y\\ \Sigma I_x I_y & \Sigma I_y^2\end{bmatrix} \begin{bmatrix}u\\v\end{bmatrix}$

SUM OVER SMALL NEIGHBOURHOOD

λ_1, λ_2 = EIGENVALUES OF M

EDGE RESPONSE $R = \lambda_1\lambda_2 - k(\lambda_1+\lambda_2)^2$ K = EMPIRICALLY DERIVED CONST.

SHI-TOMASI: BOTH EIGENVALUES ABOVE A THRESHOLD

AUTOMATIC SCALE DETECTION USING LOG - LAPLACIAN OF GAUSSIAN

CONVOLVE LOG = 2ND DERIV OF GAUSSIAN OF DIFFERENT SIZES WITH IMAGE

AT ANY X,Y, THE KERNEL OF MATCHING SIZE GIVES LARGEST RESPONSE

SIFT DESCRIPTOR = SCALE INVARIANT FEATURE TRANSFORM

IMAGE IS CONVOLVED W/ GAUSSIANS OF INCREASING SIZE
SUBTRACT ADJACENT RESULTS FOR DOG = DIFF. OF GAUSSIAN
FIND A CENTRAL FEATURE AND ITS SIZE
GET THE GRADIENTS IN A LOCAL WINDOW AROUND FEATURE
USE HISTOGRAM OF GRADIENTS TO CHARACTERISE ORIENTATION OF FEATURE
- CAN ROTATE THIS TO ALIGN WITH FEATURES IN TARGET IMAGE
- INVARIANT TO INTENSITY CHANGES, SCALE, ORIENTATION

SIZE OF GAUSSIAN KERNEL, $K = 2 \cdot \text{ceil}(3\sigma) - 1$

AUTOMATIC SCALE DETECTION - AT FEATURES CONVOLVE W/ LOG (LAPLACIAN OF GAUSSIAN)
- HIGHEST RESPONSE AT LOG OF FEATURE

FEATURE MATCHING - COMPARE FEATURES TO NEAREST NEIGHBOR

REQUIRE $D_1/D_2 < 0.8$ D_1, D_2 = DISTANCE TO NEAREST NEIGHBOURS

TRACKING - ASSUME FEATURES - INTENSITY IS CONSTANT
- MOVE UNIFORMLY
(SPATIAL COHERENCE) → - MOVE WITH OTHER LOCAL FEATURES

LUCAS-KANADE ALGORITHM = ESTIMATE SPATIAL DISPLACEMENT, HOMOGENEOUS

WE WANT CONSTANCY FRAME-TO-FRAME ⇒ $\frac{\partial I}{\partial x} + \frac{\partial I}{\partial y} + \frac{\partial I}{\partial T} = 0$

⇒ $[I_x\ I_y]\begin{bmatrix}u\\v\end{bmatrix} = -I_T$ (2 UNKNOWNS) I_x, I_y, I_T = GRADIENTS IN X, Y, T

APPLY THIS TO AN N×N PATCH $\begin{bmatrix}I_x(p_1) & I_y(p_1)\\ I_x(p_2) & I_y(p_2)\\ \vdots & \vdots \\ I_x(p_N) & I_y(p_N)\end{bmatrix}\begin{bmatrix}u\\v\end{bmatrix} = \begin{bmatrix}I_T(p_1)\\I_T(p_2)\\ \vdots\\ I_T(p_N)\end{bmatrix}$

(N NOT TOO SMALL DUE TO NOISE)

$\begin{bmatrix}\Sigma I_x^2 & \Sigma I_x I_y\\ \Sigma I_x I_y & \Sigma I_y^2\end{bmatrix}\begin{bmatrix}u\\v\end{bmatrix} = -\begin{bmatrix}\Sigma I_x I_T\\ \Sigma I_y I_T\end{bmatrix}$ minimize $\|A d - B\|^2$

$(A^T A)d = A^T B$ $A^T A$ SHOULD BE INVERTIBLE, WELL CONDITIONED
⇒ λ_1, λ_2 OF $A^T A$ SHOULD BE LARGER THAN THRESHOLD

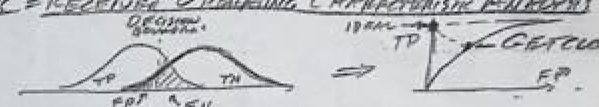
WORKS WITH SHI-TOMASI = LUCAS-KANADE-TOMASI = LKT

LK ASSUMES SMALL DISPLACEMENTS, FOR LARGE OFFSETS, SCALE DOWN THE IMAGE

KALMAN FILTER = ESTIMATES NEW LOCATIONS OF FEATURES, BASED ON PAST

INITIAL STATE → PREDICTION → (OBSERVATION) → CORRECTION

ROC = RECEIVER OPERATING CHARACTERISTIC ANALYSIS



3D RECONSTRUCTION $X_I = \frac{f}{Z_c} X_c$, $Y_I = \frac{f}{Z_c} Y_c$

X_I, Y_I = IMAGE COORDS
X_c, Y_c, Z = CAMERA COORDS
f = DISTANCE CAM → IMAGE PLANE

HOMOGENEOUS COORDS u, v, w: $X_I = \frac{u}{w}$ $Y_I = \frac{v}{w}$

CALIBRATION MATRIX $\begin{bmatrix}u\\v\\w\end{bmatrix} = \begin{bmatrix}\alpha_x & \text{skew} & X_0\\ 0 & \alpha_y & Y_0\\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}X_w\\Y_w\\Z_w\end{bmatrix}$ $K = \begin{bmatrix}\alpha_x & s & X_0\\ 0 & \alpha_y & Y_0\\ 0 & 0 & 1\end{bmatrix}$

CAMERA TRANSFORMATION MATRIX $\begin{bmatrix}X_w\\Y_w\\Z_w\end{bmatrix} = R\begin{bmatrix}X_c\\Y_c\\Z_c\end{bmatrix} + \begin{bmatrix}T_x\\T_y\\T_z\end{bmatrix}$ ← T = CAMERA POS
R = ROTATION MATRIX

ORTHOGRAPHIC PROJECTION, Z HAS NO EFFECT, $X_0 = Y_0 = 0$ (IN K), PARALLEL STAYS PARALLEL

STEREO CAMS - A POINT SEEN BY CAM C AT X, WILL APPEAR TO CAM C' AT X' ON EPIPOLAR LINE

C & C' APPEAR TO EACH OTHER AT EPIPOLAR POINTS

FUNDAMENTAL MATRIX F $x'^T F x = 0$ $[a]_x = \begin{bmatrix}0 & -a_z & a_y\\ a_z & 0 & -a_x\\ -a_y & a_x & 0\end{bmatrix}$

ESSENTIAL MATRIX $E = [T]_x R = (K')^T F K$ (T = TRANSFORMS FROM C TO C'; CALIBRATION OF C'; CALIBRATION OF C)

RECTIFIED CAMERAS DIFFER ONLY BY X OFFSET

TEXTURE MATCHING - MAKE CO-OCCURRENCE MATRICES
- NORMALIZE TO PROBABILITIES, P
- MAKE VECTORS OF RESULTS, EG ENTROPY VECTOR, ENERGY VECTOR

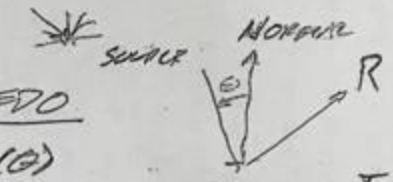
ENERGY $= \sum_{a,b} P^2(a,b)$ ENTROPY $= -\sum_{a,b} P(a,b)\log_2(P(a,b))$ CONTRAST $= \sum_{a,b} (a-b)^2 P(a,b)$ HOMOG $= \sum_{a,b} \frac{P(a,b)}{1+|a-b|}$

STEREO MATCHING

DEPTH $= d = \frac{ff'}{x-x'}$ f = FOCAL LENGTH, $C-C'$ = DISTANCE BETWEEN CAM. CENTERS

$x-x'$ = DISPARITY = PIXEL OFFSET

LAMBERT = COMPLETELY DIFFUSE



ρ = REFLECTANCE FACTOR = ALBEDO

LAMBERT COSINE LAW $R \propto S\cos(\theta)$

NORMAL N IS CROSS PRODUCT OF SLOPE VECTORS $N = [p, q, 1]^T$ (INWARD)

PHOTOMETRIC STEREO / SHAPE FROM SHADING

$R(p,q) = \rho \frac{n \cdot s}{|n||s|}$

RUNS BECAUSE CONSTANT INTENSITIES ON CONIC SECTIONS. USE NEED THREE IMAGES TO RECOVER N FROM INTENSITIES.

- ASSUMES UNIFORM LIGHTS FAR FROM OBJECT, CAMERA FAR FROM LAMBERTIAN SURFACE (NO REFLECTANCE)

ALSO SHADE FROM MOTION, TEXTURE, FOCUS/DEFOCUS, ZOOMING, CONTOURS

MOTION VECTORS / OPTICAL FLOW

$\frac{df}{dx}u + \frac{df}{dy}v + \frac{df}{dt} = 0$ ESTIMATE MOTION u, v $u = \frac{dx}{dt}$ $v = \frac{dy}{dt}$

WE CAN MEASURE $\frac{df}{dx}, \frac{df}{dy}, \frac{df}{dt}$ FROM IMAGES, BUT ONLY 1 EQUATION, 2 UNKNOWNS

ASSUME NEIGHBORING PIXELS HAVE SAME MOTION FOR ADDITIONAL CONSTRAINTS TO SOLVE FOR u, v.

APERTURE PROBLEM = MOTION IS HARD TO ESTIMATE WHEN WE ONLY OBSERVE PART

CAN'T SEE MOTION OF AN EDGE ALONG ITS EDGE, ONLY PERPENDICULAR.

FOCUS OF EXPANSION = THE POINT AT INFINITY WHERE AN OBJECT APPEARS TO COME FROM, DUE TO IT'S MOTION OR CAM MOTION.

TIME TO ADJACENCY = TIME UNTIL MOVING OBJECT HITS US = $\frac{\text{DEPTH}}{\text{VELOCITY}}$

CONVOLUTIONAL LAYER

INPUT HAS THREE DIM (X, Y, FEATURES_IN (EG. RGB)

CONV LAYER HAS FOUR DIMS (WIDTH, HEIGHT, FEATURES_IN, FEATURES_CONV)

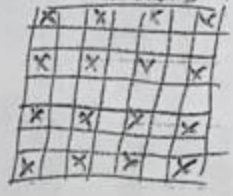
FEATURE MAP HAS THREE DIMS (X, Y, FEATURES_CONV)

- CONVOLUTIONAL LAYERS AREN'T FULLY CONNECTED TO THE INPUT, JUST A LOCALIZED SECTION

DILATION

4×4

DILATION=1



CAN REUSE THE SAME KERNEL REPEATEDLY APPLIED ACROSS THE WHOLE IMAGE

STRIDE = 2 ⇒ SHIFT THE KERNEL 2 PIXELS AT EACH STEP

POOLING = DOWNSIZE A LAYER, (EG FACTOR OF 2)

MAX POOLING = SELECT MAX VALUE IN EACH GROUP FOR WHERE