OPP
A

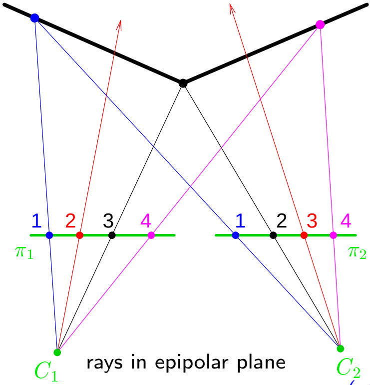
PRA HA
PRA GUE
PRA GA
PRA G

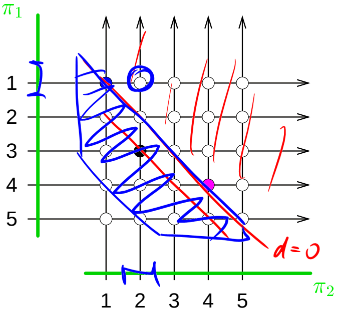
EVROPSKÁ UNIE

OPPA European Social Fund
Prague & EU: We invest in your future.

▶Matching Table

Based on the observation on mutual exclusion we expect each pixel to match at most once.





matching table T

$(3 \cdot 10^3)^3$

disparity (search) range

matching table

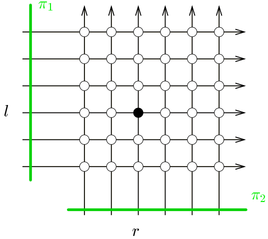
- rows and columns represent optical rays
- nodes: possible correspondence pairs
- full nodes: correspondences
- numerical values associated with nodes: descriptor similarities

see next

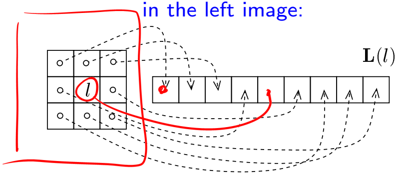
▶Constructing A Suitable Image Similarity

- let $p_i = (l, r)$ and $\mathbf{L}(l)$, $\mathbf{R}(r)$ be (left, right) image descriptors (vectors) constructed from local image neighborhood windows

in matching table T:



in the left image:



SSD, SAD

- a natural descriptor similarity is dis$\mathrm{sim}(l, r) = \dfrac{\|\mathbf{L}(l) - \mathbf{R}(r)\|^2}{\sigma_I^2(l, r)}$ ←
- σ_I^2 – the difference scale; a suitable (plug-in) estimate is $\frac{1}{2}\left[s^2\big(\mathbf{L}(l)\big) + s^2\big(\mathbf{R}(r)\big)\right]$, giving

$$\text{dis}\,\mathrm{sim}(l, r) = 1 - \underbrace{\frac{2\, s\big(\mathbf{L}(l), \mathbf{R}(r)\big)}{s^2\big(\mathbf{L}(l)\big) + s^2\big(\mathbf{R}(r)\big)}}_{\rho\left(\mathbf{L}(l), \mathbf{R}(r)\right)} \qquad (30)$$

$s^2(\cdot)$ is sample (co-)variance

- ρ – MNCC – Moravec's Normalized Cross-Correlation [Moravec 1977]

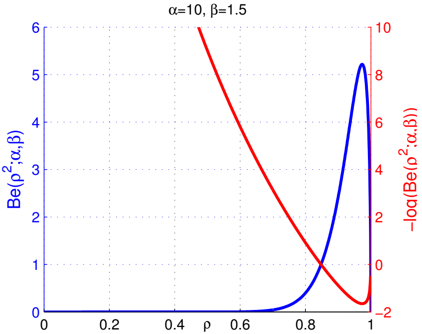
$$\rho^2 \in [0, 1], \qquad \mathrm{sign}\,\rho \sim \text{'phase'}$$

cont'd

- we choose some probability distribution on $[0,1]$, e.g. Beta distribution

$$p_1\big(\mathrm{sim}(l,r)\big) = \frac{1}{B(\alpha,\beta)}\,\rho^{2(\alpha-1)}(1-\rho^2)^{\beta-1}$$

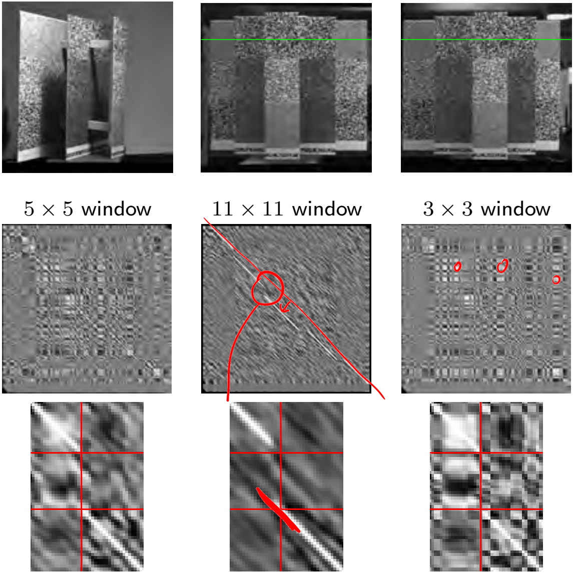
- note that uniform distribution is obtained for $\alpha = \beta = 1$



- the mode is at $\sqrt{\frac{\alpha-1}{\alpha+\beta-2}} \approx 0.9733$ for $\alpha = 10,\ \beta = 1.5$
- if we chose $\beta = 1$ then the mode was at $\rho = 1$
- perfect similarity is 'suspicious' (depends on expected camera noise level)
- from now on we will work with

$$V_1\big(\mathrm{sim}(l,r)\big) = -\log p_1\big(\mathrm{sim}(l,r)\big) \tag{31}$$

How A Scene Looks in The Filled-In Similarity Table



scene | left image | right image

5×5 window | 11×11 window | 3×3 window

a good tradeoff | occlusion artefacts | undiscrimiable

- MNCC ρ used ($\alpha = 1.5$, $\beta = 1$)
- high-correlation structures correspond to scene objects

constant disparity

- a diagonal in correlation table
- zero disparity is the main diagonal

depth discontinuity

- horizontal or vertical jump in correlation table

large image window

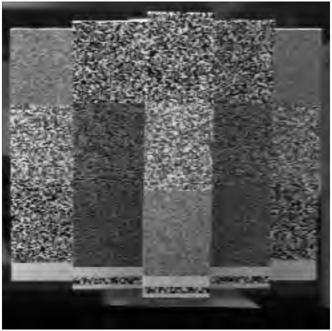
- better correlation
- worse occlusion localization

see next

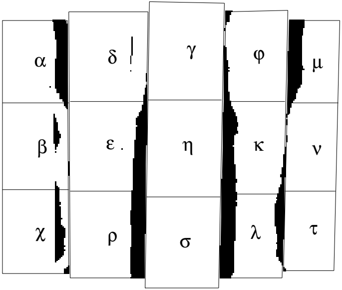
repeated texture

- horizontal and vertical block repetition

Note: Errors at Occlusion Boundaries for Large Windows



NCC, Disparity Error



- this used really large window of 25×25 px
- errors depend on the relative contrast across the occlusion boundary
- the direction of 'overlow' depends on the combination of texture contrast and edge contrast
- solutions:
  1. small windows (5×5 typically suffices)
  2. eg. 'guided filtering' methods for computing image similarity [Hosni 2011]

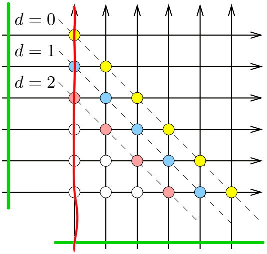
▶Marroquin's Winner Take All (WTA) Matching Algorithm

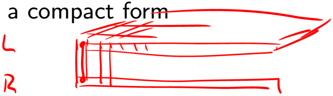
1. per left-image pixel: find the most similar right-image pixel

$$\mathrm{SAD}(l, r) = \|\mathbf{L}(l) - \mathbf{R}(r)\|_1$$

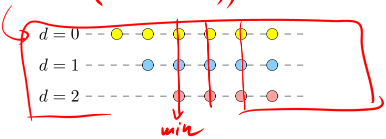
L_1 norm instead of the L_2 norm in (30); unnormalized

2. represent the dissimilarity table diagonals in a compact form



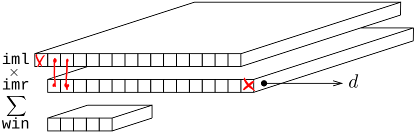


s = abs(iml - imr);



s= abs(iml(:,2:end) - imr(:,1:end-1))

3. use the 'image sliding aggregation algorithm'

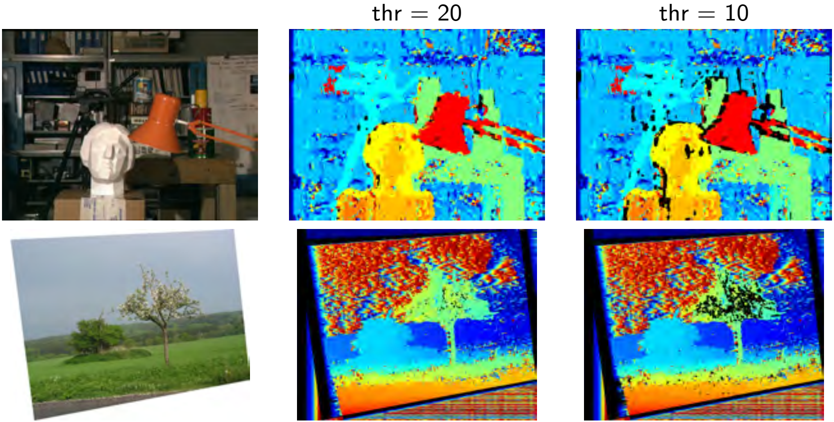


4. threshold results by maximal allowed dissimilarity

The Matlab Code for WTA

```
function dmap = marroquin(iml,imr,disparityRange)
%       iml, imr - rectified gray-scale images
% disparityRange - non-negative disparity range

% (c) Radim Sara (sara@cmp.felk.cvut.cz) FEE CTU Prague, 10 Dec 12

 thr = 20;             % bad match rejection threshold
 r = 2;
 winsize = 2*r+[1 1];  % 5x5 window (neighborhood)

 % the size of each local patch; it is N=(2r+1)^2 except for boundary pixels
 N = boxing(ones(size(iml)), winsize);

 % computing dissimilarity per pixel (unscaled SAD)
 for d = 0:disparityRange                        % cycle over all disparities
  slice = abs(imr(:,1:end-d) - iml(:,d+1:end));  % pixelwise dissimilarity
  V(:,d+1:end,d+1) = boxing(slice, winsize)./N;  % window aggregation
 end

 % collect winners, threshold, and output disparity map
 [cmap,dmap] = min(V,[],3);
 dmap(cmap > thr) = NaN;    % mask-out high dissimilarity pixels
end

function c = boxing(im, wsz)
 % if the mex is not found, run this slow version:
 c = conv2(ones(1,wsz(1)), ones(wsz(2),1), im, 'same');
end
```

WTA: Some Results



- results are bad
- false matches in textureless image regions and on repetitive structures (book shelf)
- a more restrictive threshold (thr=10) does not work as expected
- we searched the true disparity range, results get worse if the range is set wider
- chief failure reasons:
    - unnormalized image dissimilarity does not work well
    - no occlusion model

▶Negative Log-Likelihood of Observed Images

- given matching M what is the likelihood of observed data D?
- we need the ability 'not to match'
- matches are pairs $p_i = (l_i, r_i), \quad i = 1, \ldots, n$
- we will mask-out some matches by a binary label $\lambda \in \{\mathrm{e}, \mathrm{m}\}$ excluded, matched
- labeled matching is a set

$$M = \Big\{ \big(p_1, \lambda(p_1)\big), \big(p_2, \lambda(p_2)\big), \ldots, \big(p_n, \lambda(p_n)\big) \Big\}$$

p_i are matching table pairs; there are no more than n in the table T

The negative log-likelihood is then the likelihood of data D given labeled matching M

$$V(D \mid M) = \sum_{p_i \in M} V\big(D(p_i) \mid \lambda(p_i)\big)$$

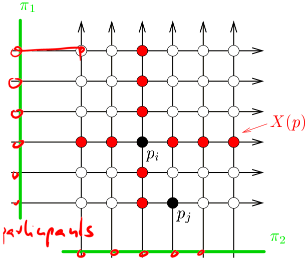
Our choice:

$$V\big(D(p_i) \mid \lambda(p_i) = \mathrm{e}\big) = V_\mathrm{e}$$ penalty for unexplained data, $V_\mathrm{e} \geq 0$

$$V\big(D(p_i) \mid \lambda(p_i) = \mathrm{m}\big) = V_1\big(D(l, r)\big)$$ probability of match $p_i = (l, r)$ from (31)

- the $V\big(D(p_i) \mid \lambda(p_i) = \mathrm{e}\big)$ could also be a non-uniform distribution but the extra effort does not pay off

▶Maximum Likelihood (ML) Matching



Uniqueness constraint: Each point in the left image matches at most once and vice versa.

A node set of T that follows the uniqueness constraint is called matching in graph theory

A set of pairs $M = \{p_i\}_{i=1}^{n}$, $p_i \in T$ is a matching iff

$$\forall p_i, p_j \in M, i \neq j :\ p_j \notin X(p_i).$$

The $X(p)$ is called the X-zone of p and it defines dependencies

- ML matching will observe the uniqueness constraint only
- epipolar lines are independent wrt uniqueness constraint
- we can solve the problem per image lines i independently:

⊛ H4; 2pt: How many are there: (1) binary partitionings of T, (2) maximal matchings in T; prove the results.

Jan 3

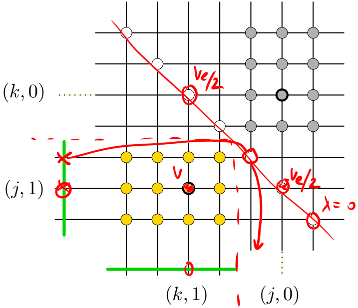
$$M^* = \mathop{\arg\min}_{M\in\mathcal{M}} \sum_{p\in M} V\big(D(p) \mid \lambda(p)\big) = \mathop{\arg\min}_{M\in\mathcal{M}} \Big(\underbrace{|M|_{\mathrm{e}} \cdot V_{\mathrm{e}}}_{\text{unexplained pixels}} + \underbrace{\sum_{p\in M:\ \lambda(p)=\mathrm{m}} V\big(D(p) \mid \lambda(p)=\mathrm{m}\big)}_{\text{matching likelihood proper}} \Big)$$

$\mathcal{M}$ – set of all perfect labeled matchings, $|M|_{\mathrm{e}}$ – number of pairs with $\lambda = \mathrm{e}$ in M, $|M|_{\mathrm{e}} \leq n$
perfect = every table row (column) contains exactly 1 match

- the total number of individual terms in the sum is n (which is fixed)

▶'Programming' The ML Matching Algorithm

- we restrict ourselves to a single (rectified) image line and reduce the problem to min-cost perfect matching
- extend every matching table pair $p \in T$, $p = (j, k)$ to 4 combinations $\big((j, s_j), (k, s_k)\big)$, $s_j \in \{0, 1\}$ and $s_k \in \{0, 1\}$ selects/rejects pixels for matching unlike λ selecting matches
- binary label $m_{jk} = 1$ then means that (j, s_j) matches (k, s_k)



🟡 $V_{jk} = V(D(j, k) \mid \lambda_{jk} = \mathrm{m})$ ● $V_{jk} = 0$

○ $V_{jk} = \frac{1}{2} V_{\mathrm{e}}$ + $V_{jk} = \infty$

- each $(j, 1)$ either matches some $(k, 1)$ or it 'matches' $(j, 0)$
- each $(k, 1)$ either matches some $(j, 1)$ or $(k, 0)$
- if M is maximal in the yellow quadrant then there will be n auxiliary 'matches' in the gray quadrant
- otherwise every empty line in the yellow quadrant induces an empty column in the quadrant, the cost is $2 \cdot \frac{1}{2} V_{\mathrm{e}} = V_{\mathrm{e}}$

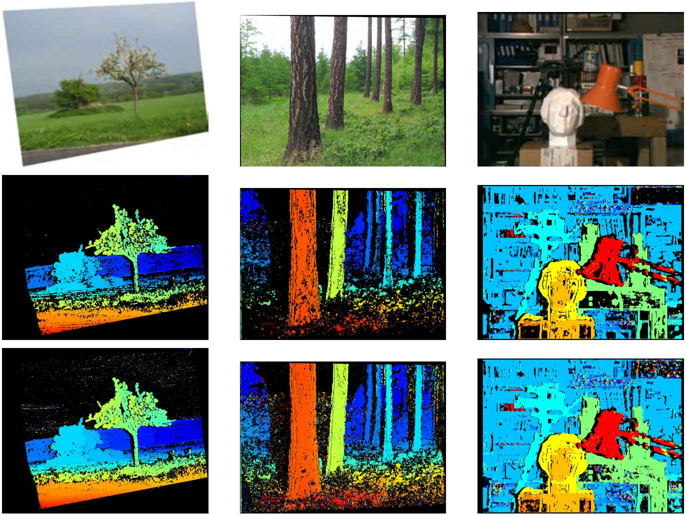
- our problem becomes minimum-cost perfect matching in an $(m + n) \times (m + n)$ table

$$M^+ = \arg\min_{M} \sum_{j,k} V_{jk} \cdot m_{jk}, \quad \sum_{k} m_{jk} = 1 \text{ for every } j, \quad \sum_{j} m_{jk} = 1 \text{ for every } k$$

$m_{jk} \in \{0,1\}$

- we collect our matches M^* in the yellow quadrant

Some Results for the ML Matching



- unlike the WTA we can efficiently control the density/accuracy tradeoff
- middle row: V_e set to error rate of 3% (and 61% density is achieved)
- bottom row: V_e set to density of 76% (and 4.3% error rate is achieved)

holes are black

Some Notes on ML Matching

- an algorithm for maximum weighted bipartite matching can be used as well, with $V \mapsto -V$
- maximum weighted bipartite matching = maximum weighted assignment problem

by eg. Hungarian Algorithm

Idea?: This looks simpler: Run matching with $V_{\mathrm{e}} = 0$ and then threshold the result to remove bad matches.

Ex: $V_{\mathrm{e}} = 8$

thresholding

8	3	**9**
10	**6**	9
7	1	8

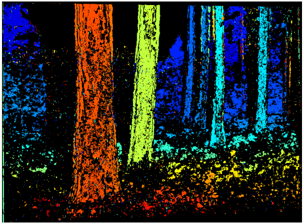
$V = 9 + 2 \cdot 8 = 25$

our ML matching

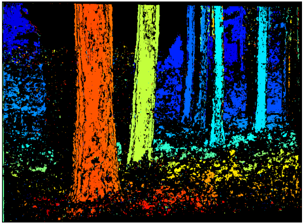
8	3	**9**
10	6	9
7	1	8

$V = 9 + 10 + 8 = 27$

- our matching gives a better cost, also greater cardinality (density)
- the idea was not good!



thresholding



our ML

A Stronger Model Needed

- notice many small isolated errors in the ML matching
- we need a continuity model
- does human stereopsis teach us something?

Potential models for M

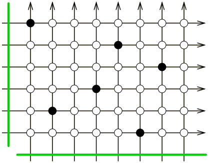
1. Monotonicity (ie. ordering preserved):

$$\text{For all } (i,j) \in M, (k,l) \in M, \quad k > i \Rightarrow l > j$$

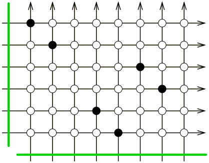
Notation: $(i,j) \in M$ or $j = M(i)$ – left-image pixel i matches right-image pixel j.

2. Coherence [Prazdny 85]

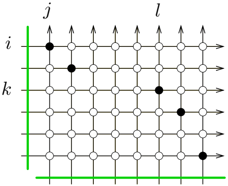
"the world is made of objects each occupying a well defined 3D volume"



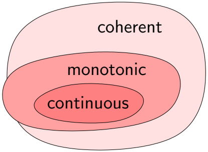
non-monotonic
incoherent



non-monotonic
coherent



monotonic
coherent

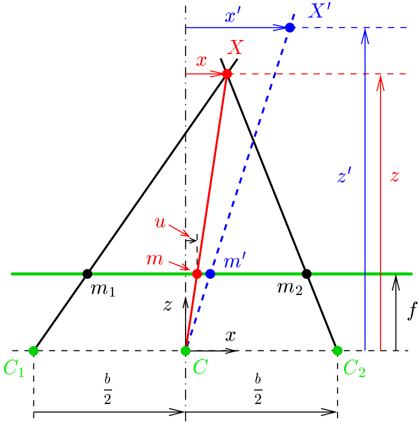


model 'strength'

▶An Auxiliary Construct: Cyclopean Camera

Cyclopean coordinate u from the psychophysiology of vision [Julesz 1971]

$$\text{new: } u = f\,\frac{x}{z}, \quad \text{known: } d = f\,\frac{b}{z}, \quad x = \frac{b}{d}\,\frac{u_1 + u_2}{2} \quad \Rightarrow \quad u = \frac{u_1 + u_2}{2}$$



Disparity gradient [Pollard, Mayhew, Frisby 1985]

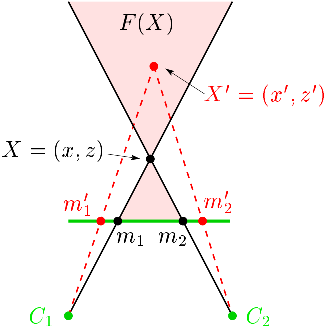
$$DG = \frac{|d - d'|}{|u - u'|} = \frac{\left|bf\left(\frac{1}{z} - \frac{1}{z'}\right)\right|}{\left|f\left(\frac{x}{z} - \frac{x'}{z'}\right)\right|} = $$

$$= b\,\frac{|z' - z|}{|xz' - x'z|}$$

- human stereovision fails to perceive a continuous surface when disparity gradient exceeds a limit

▶Forbidden Zone and The Ordering Constraint

Forbidden zone $F(X)$: $DG > k$ with boundary $b\,(z' - z) = \pm k\,(xz' - x'z)$



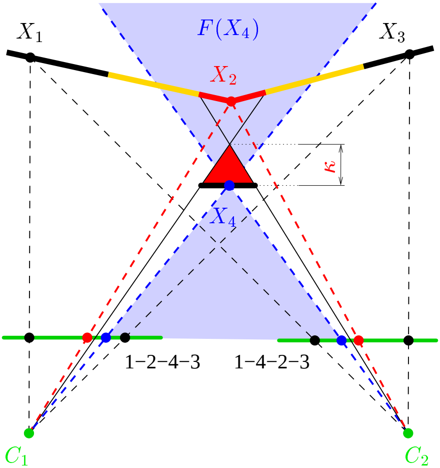
- boundary: a pair of lines in the $x - z$ plane

a degenerate conic

- point $x = x'$, $z = z'$ lies on the boundary
- coincides with optical rays for $k = 2$
- small k means wide F

- disparity gradient limit is exceeded when $X' \in F(X)$
- symmetry: $X' \in F(X) \Leftrightarrow X \in F(X')$
- **Obs:** X' and X swap their order in the other image when $X' \in F(X)$ $k = 2$
- real scenes often preserve ordering
- thin and close objects violate ordering see next

Ordering and Critical Distance κ



- object (thick):
    - black – binocularly visible
    - yellow – half-occluded
    - red – ordering violated wrt foreground
- solid red zone of depth κ:
    - spatial points visible in neither camera
    - bounded by the foreground object

Ordering is violated iff both X_i, X_j s.t. $X_i \in F(X_j)$ are visible in both cameras.

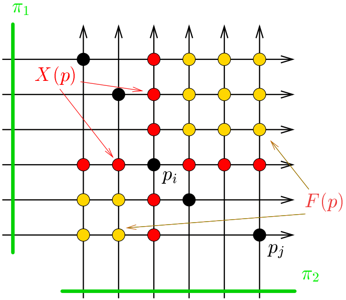
eg. X_2, X_4

- ordering is preserved in scenes where critical distances κ are not exceeded, ie. when 'the red background hides in the solid red zone'

Thinner objects and/or wider baseline require flatter scenes to preserve ordering.

The X-zone and the F-zone in Matching Table T

- these are necessary and sufficient conditions for uniqueness and monotonicity



$$p_j \notin X(p_i), \quad p_j \notin F(p_i)$$

- **Uniqueness Constraint:**

A set of pairs $M = \{p_i\}_{i=1}^{N}$, $p_i \in T$ is a matching iff

$$\forall p_i, p_j \in M, i \neq j : \; p_j \notin X(p_i).$$

- **Ordering Constraint:**

Matching M is monotonic iff

$$\forall p_i, p_j \in M : \; p_j \notin F(p_i).$$

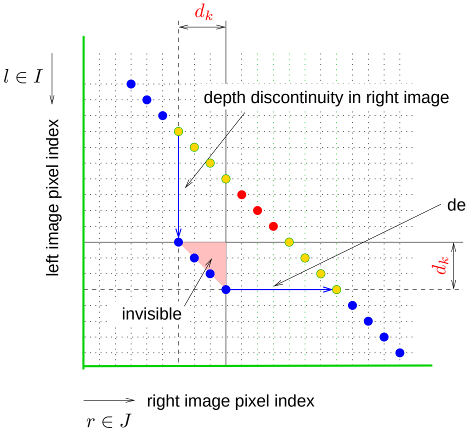
  - ordering constraint: matched points form a monotonic set in both images
  - ordering is a powerful constraint: monotonic matchings $O(4^N) \ll O(N!)$ all matchings in $N \times N$ table

⊛ 2: how many are there maximal monotonic matchings?

- uniqueness constraint is a basic occlusion model
- ordering constraint is a weak continuity model and partly also an occlusion model

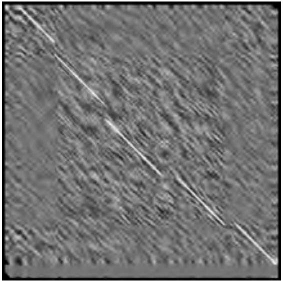
Understanding Matching Table

- this is essentially the picture from Slide 178



- binocularly visible foreground points
- binocularly visible background pts violating ordering
- monocularly visible points

d_k critical disparity



Bayesian Decision Task for Matching

Idea: $L(d, M)$ – decision cost (loss) d – our decision (matching) M – true correspondences

Choice: $L(d, M) : \begin{cases} \text{if } d = M & \text{then } L(d, M) = 0 \\ \text{if } d \neq M & \text{then } L(d, M) = 1 \end{cases}$ i.e. $L(d, M) = [d \neq M]$

Bayesian Loss

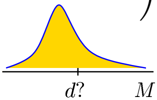
$$L(d \mid D) = \sum_{M \in \mathcal{M}} p(M \mid D)\, L(d, M)$$

$\mathcal{M}$ – the set of all matchings $D = \{I_L, I_R\}$ – data

Solution for the best decision d

$$d^* = \arg\min_{d} \sum_{M \in \mathcal{M}} p(M \mid D)\,(1 - [d = M]) = \arg\min_{d} \left(1 - \sum_{M \in \mathcal{M}} p(M \mid D)[d = M]\right) =$$

$$= \arg\max_{d} \sum_{M \in \mathcal{M}} p(M \mid D)\,[d = M] = \arg\max_{M} p(M \mid D) =$$



$$= \arg\min_{M}(-\log p(M \mid D)) \overset{\text{def}}{=} \arg\min_{M} V(M \mid D) = \arg\min_{M \in \mathcal{M}} \Big(\underbrace{V(D \mid M)}_{\text{likelihood}} + \underbrace{V(M)}_{\text{prior}}\Big)$$

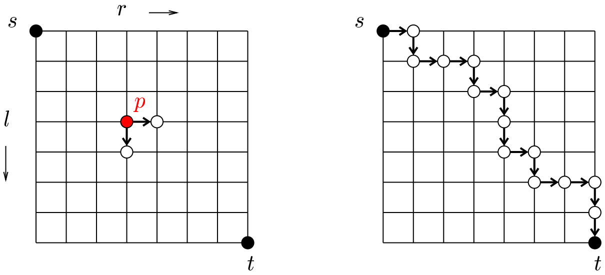
- this is Maximum Aposteriori Probability (MAP) estimate
  - other loss functions result in different solutions
  - our choice of $L(d, M)$ looks oversimple but it results in algorithmically tractable problems

Constructing The Prior Model Term $V(M)$

- the prior $V(M)$ should capture
  1. uniqueness
  2. ordering
  3. coherence

$$M^* = \arg\min_{M \in \mathcal{M}} \Big(V(D \mid M) + V(M) \Big)$$

- we need a suitable representation to encode $V(M)$
  - Every $p = (l, r)$ of the $|I| \times |J|$ matching table T (except for the last row and column) receives two succesors $(l+1, r)$ and $(l, r+1)$

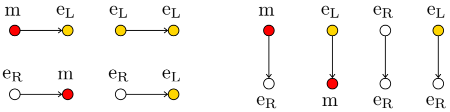


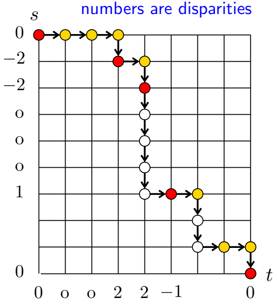
  - this gives an acyclic directed graph $\mathcal{G}$ optimal paths in acyclic graphs are an easier problem
  - the set of s-t paths starting in s and ending in t will represent the set of matchings
  - all such s-t paths have **equal** length $n = |I| + |J| - 1$

all prospective matchings will have the same number of terms in $V(D \mid M)$ and in $V(M)$

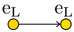
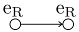
Endowing s-t Paths with Useful Properties

- introduce node labels $\Lambda = \{\mathrm{m}, \mathrm{e_L}, \mathrm{e_R}\}$ matched, left-excluded, right-excluded
- s-t path neighbors are allowed only some label combinations:



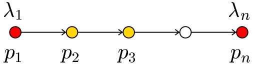

numbers are disparities

Observations

- no two neighbors have label m
- in each labeled s-t path there is at most one transition:
  1. $\mathrm{m} \to \mathrm{e_L}$ or $\mathrm{e_R} \to \mathrm{m}$ per matching table row,
  2. $\mathrm{m} \to \mathrm{e_R}$ or $\mathrm{e_L} \to \mathrm{m}$ per matching table column
- pairs labeled m on every s-t path satisfy uniqueness and ordering constraints
- transitions $\mathrm{e_L} \to \mathrm{e_R}$ or $\mathrm{e_R} \to \mathrm{e_L}$ along an s-t path allow skipping a contiguous segment in either or in both images this models half occlusion and mutual occlusion
- disparity change is the number of edges  or 
- a given monotonic matching can be traversed by one or more s-t paths

Labeled s-t paths

$$P = \big((p_1, \lambda_1),\ (p_2, \lambda_2), \dots, (p_n, \lambda_n)\big)$$

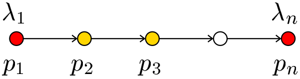


The Structure of The Prior Model $V(P)$ Gives a MC Recognition Problem

ideas:

- we choose energy of path P dependent on its labeling only
- we choose additive penalty per transition $\mathrm{e_L} \to \mathrm{e_L}$, $\mathrm{e_R} \to \mathrm{e_R}$, and $\mathrm{e_L} \to \mathrm{e_R}$, $\mathrm{e_R} \to \mathrm{e_L}$
- no penalty for $\mathrm{m} \to \mathrm{e_L}$, $\mathrm{m} \to \mathrm{e_R}$

Employing Markovianity



$$V(P) = V(\lambda_n, \lambda_{n-1}, \ldots, \lambda_1) = V(\lambda_n \mid \lambda_{n-1}, \ldots, \lambda_1) + V(\lambda_{n-1}, \ldots, \lambda_1) =$$

$$= V(\lambda_n \mid \lambda_{n-1}) + V(\lambda_{n-1}, \ldots, \lambda_1) = V(\lambda_1) + \sum_{i=2}^{n} V(\lambda_i \mid \lambda_{i-1})$$

The matching problem is then a decision over labeled s-t paths $P \in \mathcal{P}$:

$$P^* = \arg\min_{P \in \mathcal{P}} \left\{ V_{p_1}(D \mid \lambda_1) + V(\lambda_1) + \sum_{i=2}^{n} \Bigl[V_{p_i}(D \mid \lambda_i) + V(\lambda_i \mid \lambda_{i-1}) \Bigr] \right\} \qquad (32)$$

- the data likelihood term $V_{p_i}(D \mid \lambda_i)$ is the same as in (31) on Slide 164
- note that one can add/subtract a fixed term from any of the functions V_p, V in (32)

A Choice of $V(\lambda_i \mid \lambda_{i-1})$

- A natural requirement: symmetry of probability $p(\lambda_i, \lambda_{i-1}) = e^{-V(\lambda_i,\, \lambda_{i-1})}$

$p(\lambda_i, \lambda_{i-1})$		λ_i		
		m	$\mathrm{e_L}$	$\mathrm{e_R}$
λ_{i-1}	m	0	$p(\mathrm{m}, \mathrm{e})$	$p(\mathrm{m}, \mathrm{e})$
	$\mathrm{e_L}$	$p(\mathrm{m}, \mathrm{e})$	$p(\mathrm{e}, \mathrm{e})$	$p(\mathrm{e_L}, \mathrm{e_R})$
	$\mathrm{e_R}$	$p(\mathrm{m}, \mathrm{e})$	$p(\mathrm{e_L}, \mathrm{e_R})$	$p(\mathrm{e}, \mathrm{e})$

3 DOF, 1 constraint $\Rightarrow$ 2 parameters

$$\alpha_1 = \frac{p(\mathrm{e_L}, \mathrm{e_R})}{p(\mathrm{e}, \mathrm{e})} \qquad 0 \leq \alpha_1 \leq 1$$

$$\alpha_2 = \frac{p(\mathrm{m}, \mathrm{e})}{p(\mathrm{e}, \mathrm{e})} \qquad 0 < \alpha_2 \leq 1 + \alpha_1$$

- **Result** for $V(\lambda_i \mid \lambda_{i-1})$ (after subtracting common terms):

$V(\lambda_i \mid \lambda_{i-1})$		λ_i		
		m	$\mathrm{e_L}$	$\mathrm{e_R}$
λ_{i-1}	m	∞	0	0
	$\mathrm{e_L}$	$\ln \frac{1+\alpha_1+\alpha_2}{2\,\alpha_2}$	$\ln \frac{1+\alpha_1+\alpha_2}{2}$	$\ln \frac{1+\alpha_1+\alpha_2}{2\,\alpha_1}$
	$\mathrm{e_R}$	$\ln \frac{1+\alpha_1+\alpha_2}{2\,\alpha_2}$	$\ln \frac{1+\alpha_1+\alpha_2}{2\,\alpha_1}$	$\ln \frac{1+\alpha_1+\alpha_2}{2}$

by marginalization:

$$V(\mathrm{m}) = \ln \frac{1+\alpha_1+\alpha_2}{2\,\alpha_2}$$

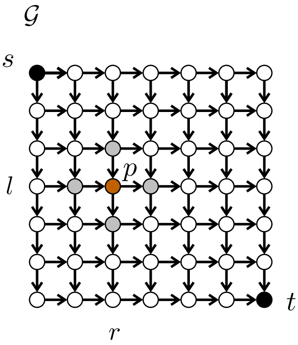
$$V(\mathrm{e_L}) = V(\mathrm{e_R}) = 0$$

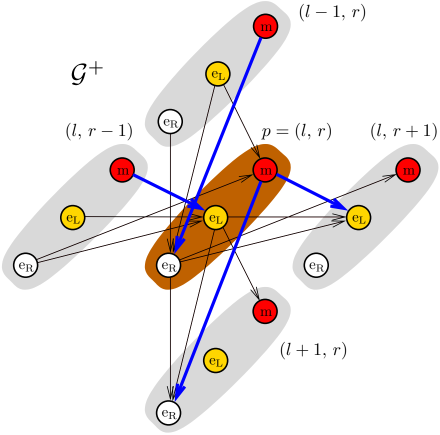
parameters

- α_1 – likelihood of mutual occlusion ($\alpha_1 = 0$ forbids mutual occlusion)
- α_2 – likelihood of irregularity ($\alpha_2 \to 0$ helps suppress small objects and holes)
- α, β – similarity model parameters (see $V_1\big(D(l,r)\big)$ on Slide 164)
- V_e – penalty for disregarded data (see $V(D(p_i) \mid \lambda(p_i) = \mathrm{e})$ on Slide 170)

‘Programming’ the Matching Algorithm: 3LDP

- given $\mathcal{G}$, construct directed graph $\mathcal{G}^+$
- triple of vertices per node of s-t path representing three hypotheses $\lambda(p)$ for $\lambda \in \Lambda$
- arcs have costs $V(\lambda_i \mid \lambda_{i-1})$, nodes have costs $V(D \mid \lambda_i)$
- orientation of $\mathcal{G}^+$ is inherited from the orientation of s-t paths
- we converted the shortest labeled-path problem to ordinary shortest path problem

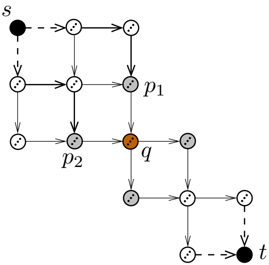




neighborhood of p; strong blue edges are of zero penalty

cont'd: Dynamic Programming on $\mathcal{G}^+$

- $\mathcal{G}^+$ is a topologically ordered directed graph
- we can use dynamic programming on $\mathcal{G}^+$



$$V^*_{s:q}(\lambda_q) = \min_{z\in\{p_1,p_2\},\lambda_z\in\Lambda} \Big\{ V^*_{s:z}(\lambda_z) + V_z(D \mid \lambda_z) + V(\lambda_q \mid \lambda_z) \Big\}$$

$V^*_{s:q}(\lambda_q)$ – cost of min-path from s to label λ_q at node q

- complexity is $O(|I| \cdot |J|)$, ie. stereo matching on $N \times N$ images needs $O(N^3)$ time
- speedup by limiting the range in which the disparities $d = l - r$ are allowed to vary

Implementation of 3LDP in a few lines of code. . .

```
#define clamp(x, mi, ma) ((x) < (mi) ? (mi) : ((x) > (ma) ? (ma) : (x)))
#define MAXi(tab,j) clamp((j)+(tab).drange[1], (tab).beg[0], (tab).end[0])
#define MINi(tab,j) clamp((j)+(tab).drange[0], (tab).beg[0], (tab).end[0])

#define ARG_MIN2(Ca, La, C0, L0, C1, L1) if ((C0) < (C1)) { Ca = C0; La = L0; } else { Ca = C1; La = L1; }

#define ARG_MIN3(Ca, La, C0, L0, C1, L1, C2, L2) \
 if ( (C0) <= MIN(C1, C2) ) { Ca = C0; La = L0; } else if ( (C1) < MIN(C0, C2) ) { Ca = C1; La = L1; } else { Ca = C2; La = L2; }
```

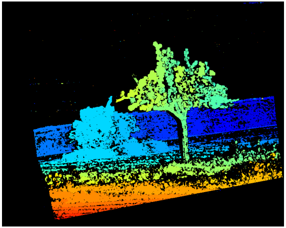
```
 void DP3LForward(MatchingTableT tab) {

  int i = tab.beg[0]; int j = tab.beg[1];
  C_m[j][i-1] = C_m[j-1][i] = MAXDOUBLE;
  C_oL[j][i-1] = C_oR[j-1][i] = 0.0;
  C_oL[j-1][i] = C_oR[j][i-1] = -penalty[0];

  for(j = tab.beg[1]; j <= tab.end[1]; j++)
   for(i = MINi(tab,j); i <= MAXi(tab,j); i++) {

     ARG_MIN2(C_m[j][i], P_m[j][i],
              C_oR[j-1][i] + penalty[2], lbl_oR,
              C_oL[j][i-1] + penalty[2], lbl_oL);
     C_m[j][i] += 1.0 - tab.MNCC[j][i];

     ARG_MIN3(C_oL[j][i], P_oL[j][i], C_m[j-1][i], lbl_m,
              C_oL[j-1][i] + penalty[0], lbl_oL,
              C_oR[j-1][i] + penalty[1], lbl_oR);
     C_oL[j][i] += penalty[3];

     ARG_MIN3(C_oR[j][i], P_oR[j][i], C_m[j][i-1], lbl_m,
              C_oR[j][i-1] + penalty[0], lbl_oR,
              C_oL[j][i-1] + penalty[1], lbl_oL);
     C_oR[j][i] += penalty[3];
   }
 }
```

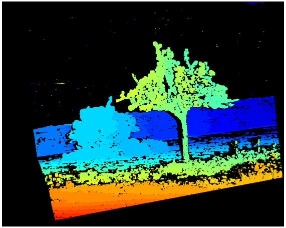
```
void DP3LReverse(double *D, MatchingTableT tab) {

 int i,j; labelT La; double Ca;
 for(i=0; i<nl; i++) D[i] = nan; /* not-a-number */

 i = tab.end[0]; j = tab.end[1];
 ARG_MIN3(Ca, La, C_m[j][i], lbl_m,
          C_oL[j][i], lbl_oL, C_oR[j][i], lbl_oR);

 while (i >= tab.beg[0] && j >= tab.beg[1] && La > 0)
  switch (La) {
   case lbl_m: D[i] = i-j;
    switch (La = P_m[j][i]) {
     case lbl_oL: i--; break;
     case lbl_oR: j--; break;
     default: Error(...);
    } break;

   case lbl_oL: La = P_oL[j][i]; j--; break;
   case lbl_oR: La = P_oR[j][i]; i--; break;
   default: Error(...);
  }
}
```

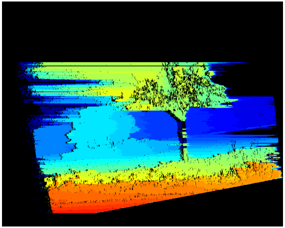
Some Results: AppleTree

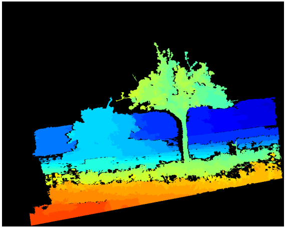

left image


right image


ML (slide 172)


3LDP (slide 186)


naïve DP [Cox et al. 1992]

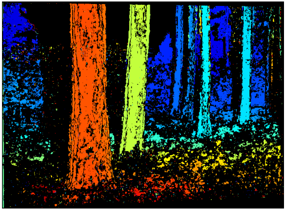

stable segmented 3LDP (see [SP])

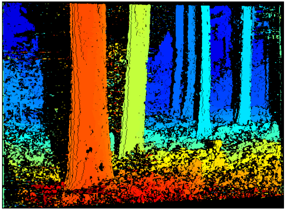
- 3LDP parameters α_i, V_e learned on Middlebury stereo data http://vision.middlebury.edu/stereo/

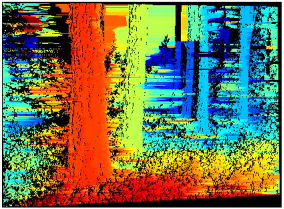
Some Results: Larch

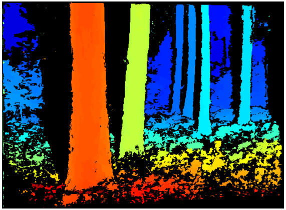

left image


right image


ML (slide 172)


3LDP (slide 186)


naïve DP


stable segmented 3LDP

- naïve DP does not model mutual occlusion
- but even 3LDP has errors in mutually occluded region
- stable segmented 3LDP has few errors in mutually occluded region since it uses a weak form of ‘image understanding’

Algorithm Comparison

Winner-Take-All (WTA)

- the ur-algorithm [Marroquin 83] no model
- dense disparity map
- $O(N^3)$ algorithm, simple but it rarely works

Maximum Likelihood (ML)

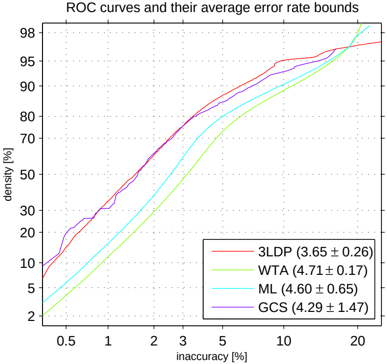
- semi-dense disparity map
- many small isolated errors
- models basic occlusion
- $O(N^3 \log(NV))$ algorithm max-flow by cost scaling

MAP with Min-Cost Labeled Path (3LDP)

- semi-dense disparity map
- models occlusion in flat, piecewise continuos scenes
- has 'illusions' if ordering does not hold
- $O(N^3)$ algorithm

Stable Segmented 3LDP

- better (fewer errors at any given density)
- $O(N^3 \log N)$ algorithm
- requires image segmentation itself a difficult task



- ROC-like curve captures the density/accuracy tradeoff
- GCS is the one used in the exercises
- more algorithms at `http://vision.middlebury.edu/stereo/` (good luck!)

Part VIII

Shape from Reflectance

mostly covered by

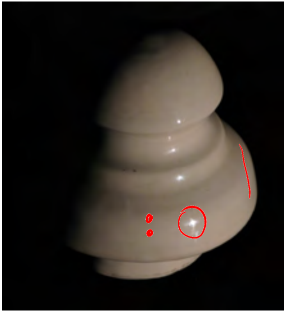
Forsyth, David A. and Ponce, Jean. *Computer Vision: A Modern Approach*. Prentice Hall 2003. Chap. 5

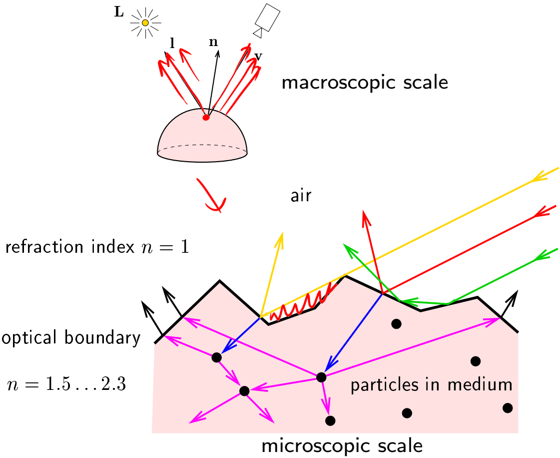
additional references

R. T. Frankot and R. Chellappa. A method for enforcing integrability in shape from shading algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10(4):439–451, July 1988.

P. N. Belhumeur, D. J. Kriegman, and A. L. Yuille. The bas-relief ambiguity. In *Proc Conf Computer Vision and Pattern Recognition*, pp. 1060–1066, 1997.

▶Basic Surface Reflectance Mechanisms





- 🔴 reflection on (rough) optical boundary
- 🟡 masking and shadowing
- 🟢 interreflection
- 🔵 refraction into the body
- 🟣 subsurface scattering
- 🔴 refraction into the air

▶Parametric Reflectance Models

Image intensity (measurement) at pixel m given by surface reflectance function R

$$J(m) = \eta\, f_{i,r}(\theta_i,\phi_i;\theta_r,\phi_r)\cdot \underbrace{\frac{\Phi_e}{4\pi\|\mathbf{L}-\mathbf{x}\|^2}}_{\sigma}\,\mathbf{n}^\top\mathbf{l} = R(\mathbf{n}), \qquad \mathbf{l} = \frac{\mathbf{L}-\mathbf{x}}{\|\mathbf{L}-\mathbf{x}\|}$$

η – sensor sensitivity for simplicity, we select $\eta = 2\pi$

$f_{i,r}()$ – bidirectional reflectance distribution function (BRDF)

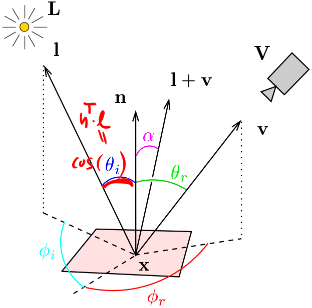
$[f_{i,r}()] = \mathrm{sr}^{-1}$ how much of irradiance in Wm^{-2} is redistributed per solid angle element

$\mathbf{L}$ – point light source position

Φ_e – radiant power of the light source, $[\Phi_e] = \mathrm{W}$

$\mathbf{n}$ – surface normal

σ – irradiance of a surfel orthogonal to incident light direction



pixel projected onto surface

Isotropic (Lambertian) reflection

[Lambert 1760]

no optical boundary

$$f_{i,r}(\theta_i,\phi_i;\theta_r,\phi_r) = \frac{\rho}{2\pi}, \qquad \rho \text{ – albedo}$$

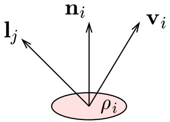
$$J(m) = \sigma\rho\cos\theta_i = \sigma\rho\,\mathbf{n}^\top\mathbf{l}$$

▶Photometric Stereo

Lambertian model (light $j \in \{1, 2, 3\}$, pixel $i \in \{1, \dots, n\}$)

$$J_{ji} = \underline{(\sigma_j\, \mathbf{l}_j)}^\top \underline{(\rho_i\, \mathbf{n}_i)} = \mathbf{s}_j^\top\, \mathbf{b}_i$$

$\mathbf{b}_i$ – scaled normals, $\mathbf{s}_j$ – scaled lights



3 independent scaled lights and n scaled normals, one per pixel (in n pixels); can be stacked in matrices:

$$\begin{array}{c} \rightarrow \\ \rightarrow \\ \rightarrow \end{array} \begin{bmatrix} J_{11} & J_{12} \\ J_{21} & J_{22} \\ \underset{\substack{\uparrow \\ \text{pixel}}}{J_{31}} & J_{32} \end{bmatrix} = \begin{bmatrix} \mathbf{s}_1^\top \mathbf{b}_1 & \mathbf{s}_1^\top \mathbf{b}_2 \\ \mathbf{s}_2^\top \mathbf{b}_1 & \mathbf{s}_2^\top \mathbf{b}_2 \\ \mathbf{s}_3^\top \mathbf{b}_1 & \mathbf{s}_3^\top \mathbf{b}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{s}_1^\top \\ \mathbf{s}_2^\top \\ \mathbf{s}_3^\top \end{bmatrix} \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 \end{bmatrix}$$

$n = 2$ pixels, 3 lights

pixel indexing i:

1	2	3	4
5	6	7	8
9	10	11	12

in general, stacked per columns:

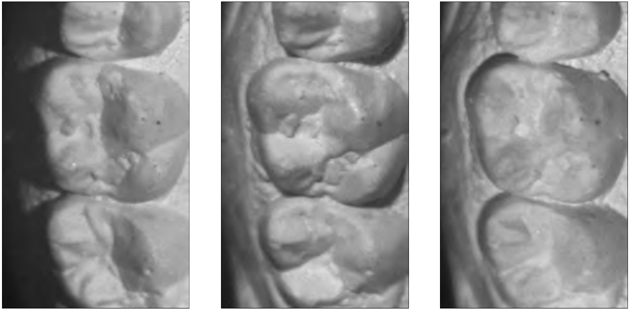
$$\mathbf{S} = [\,\mathbf{s}_1,\, \mathbf{s}_2,\, \mathbf{s}_3\,] \in \mathbb{R}^{3,3} \qquad \mathbf{B} = [\,\mathbf{b}_1,\, \mathbf{b}_2,\, \dots,\, \mathbf{b}_n] \in \mathbb{R}^{3,n}$$

Solution to Photometric Stereo

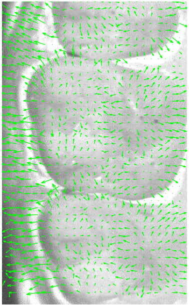
$$\mathbf{J} = \mathbf{S}^\top \mathbf{B} \quad \Rightarrow \quad \mathbf{B} = \mathbf{S}^{-\top} \mathbf{J} \qquad\qquad \mathbf{J} \in \mathbb{R}^{3,n}$$

$\rho_i = \|\mathbf{b}_i\|$ albedo map, $\qquad \mathbf{n}_i = \frac{1}{\rho_i}\, \mathbf{b}_i$ needle map

Photometric Stereo: Plaster Cast Example



input images (known lights)



needle & albedo maps

We have: 1. shape (surface normals), 2. intrinsic texture (albedo)

The shape can be represented as unit normal vectors $\mathbf{n}$ or as a gradient field (p, q):

$$\mathbf{n}(u,v) = \big(n_1(u,v),\, n_2(u,v),\, n_3(u,v)\big),$$

$$\frac{\partial z(u,v)}{\partial u} \overset{\text{def}}{=} z_u(u,v) = p(u,v) = \pm\frac{n_1(u,v)}{2n_3(u,v)^2 - 1},$$

$$\frac{\partial z(u,v)}{\partial v} \overset{\text{def}}{=} z_v(u,v) = q(u,v) = \pm\frac{n_2(u,v)}{2n_3(u,v)^2 - 1}$$

The Integration Algorithm of Frankot and Chellappa (FC)

Task: Given gradient fields $p(u,v)$, $q(u,v)$, find height function $z(u,v)$ such that z_u is close to p and z_v is close to q in the sense of a functional norm.

$$z^* = \arg\min_z Q(z), \qquad Q(z) = \iint |z_u(u,v) - p(u,v)|^2 + |z_v(u,v) - q(u,v)|^2 \, du\, dv$$

In the Fourier domain this can be written as $\mathcal{F}(z;\boldsymbol{\omega}) = \frac{1}{2\pi} \iint z(u,v) e^{-j(u\omega_u + v\omega_v)} \, du\, dv$

$$Q(z) = \iint \underbrace{|j\omega_u \, \mathcal{F}(z;\boldsymbol{\omega}) - \mathcal{F}(p;\boldsymbol{\omega})|^2 + |j\omega_v \, \mathcal{F}(z;\boldsymbol{\omega}) - \mathcal{F}(q;\boldsymbol{\omega})|^2}_{A(\mathcal{F}(z;\boldsymbol{\omega}))} \, d\boldsymbol{\omega}, \qquad \boldsymbol{\omega} = (\omega_u, \omega_v)$$

and its minimiser is from vanishing formal derivative of $A(\mathcal{F}(z;\boldsymbol{\omega}))$ wrt $\mathcal{F}(z;\boldsymbol{\omega})$

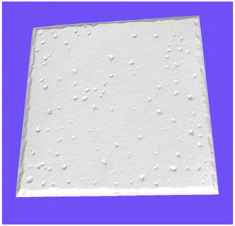
[Frankot & Chellappa 1988]

$$\mathcal{F}(z;\boldsymbol{\omega}) = -\frac{j\omega_u}{|\boldsymbol{\omega}|^2} \, \mathcal{F}(p;\boldsymbol{\omega}) - \frac{j\omega_v}{|\boldsymbol{\omega}|^2} \, \mathcal{F}(q;\boldsymbol{\omega})$$

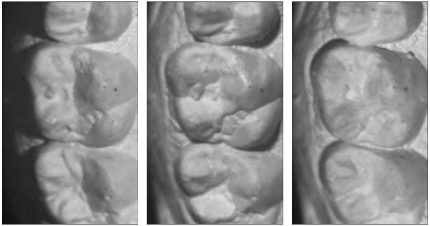
```
[m,n] = size(p);
Wu = fft2(fftshift([-1,0,1]/2),m,n); % discrete differential operator
Wv = fft2(fftshift([-1;0;1]/2),m,n);
Z = -(Wu.*fft2(p) + Wv.*fft2(q))./(abs(Wu).^2 + abs(Wv).^2 + eps);
z = real(ifft2(Z));
```

Photometric Stereo: Examples

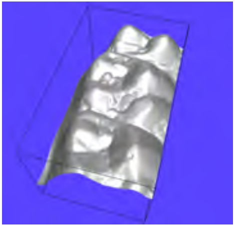
3 input images



surface



3 input images



surface

lecturing finished here

the remaining material will not be examined

- integrated by the FC algorithm from Slide 197
- bias due to interreflections can be removed

[Drew & Funt, JOSA-A 1992]

Integrability of a Vector Field

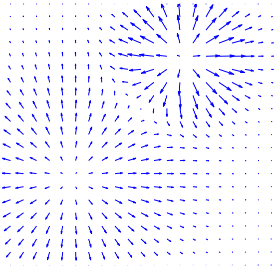
- not every vector field $p(u,v)$, $q(u,v)$ is integrable (born by a surface $z(u,v)$)
- integrability constraint

$$p_v(u,v) = q_u(u,v)$$

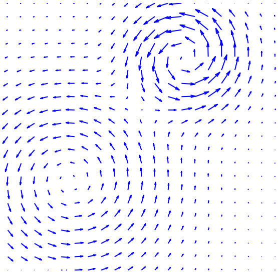
- this is because a regular surface has $\operatorname{rot} \nabla z(u,v) = 0$ irrotational gradient field

$$z_{uv}(u,v) = z_{vu}(u,v)$$

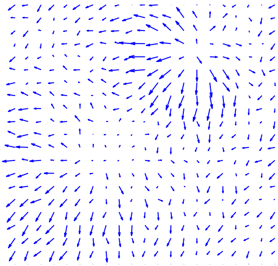
- noise causes non-integrability
- the FC algorithm finds the closest integrable surface



integrable



non-integrable



non-integrable (noisy)

Optimal Light Configurations

For n lights $\mathbf{S}$ the error $\Delta\mathbf{b} = \mathbf{S}^{-\top}\Delta\mathbf{J}$ in normal $\mathbf{b}$ due to error $\Delta\mathbf{J}$ in image is

$$\epsilon(\mathbf{S}) = E\big[\Delta\mathbf{b}^\top\Delta\mathbf{b}\big] = E\big[\Delta\mathbf{J}^\top(\mathbf{S}^\top\mathbf{S})^{-1}\Delta\mathbf{J}\big] = \sigma^2\,\mathrm{tr}\big[(\mathbf{S}\mathbf{S}^\top)^{-1}\big] \geq \frac{9\sigma^2}{n}.$$

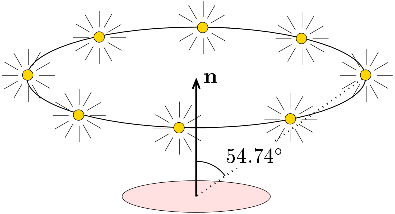
assuming pixel-independent normal camera noise $\Delta J_i \sim N(0,\sigma)$

The error ϵ is minimum if [Drbohlav & Chantler 2005]

$$\mathbf{S}\mathbf{S}^\top = \frac{n}{3}\,\mathbf{I}, \qquad \text{where} \qquad \mathbf{S} = [\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_n]$$

- either $n \geq 3$ equidistant and equiradiant lights on a circle of uniform slant of $\arctan\sqrt{2} \approx 54.74°$
- $n-1$ lights in this configuration plus a light parallel to the sum $\sum_{i=1}^{n-1}\mathbf{s}_i$
- or light matrix $\mathbf{S}$ is a concatenation of optimal solutions (each of ≥ 3 lights)

eg. 3 optimally placed $(\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3)$ + 3 lights $(\mathbf{s}_4, \mathbf{s}_5, \mathbf{s}_6) = (\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3) + \alpha$ rotated by angle α around $\mathbf{n}$



Uncalibrated Photometric Stereo

Factorization $\mathbf{J} = \mathbf{S}^\top \mathbf{B}$ [Hayakawa94]

LS solution by SVD decomposition of $\mathbf{J} = \mathbf{U}\mathbf{D}\mathbf{V}^\top$

$\mathbf{S} = \mathbf{D}_{1:3}\mathbf{U}^\top$ scaled pseudo-lights

$\mathbf{B} = (\mathbf{V}_{1:3})^\top$ scaled pseudo-normals $\mathbf{V}_{1:3}$ are columns 1–3

Ambiguity $\mathbf{J} = \mathbf{S}^\top \mathbf{B} = \underbrace{\mathbf{S}^\top \mathbf{A}^{-1}}_{\bar{\mathbf{S}}^\top} \underbrace{\mathbf{A}\mathbf{B}}_{\bar{\mathbf{B}}}, \quad \mathbf{A} \in GL(3)$ [Koenderink94]

information	ambiguity	
3+ normals $\bar{\mathbf{B}}$ known	$\lambda\mathbf{I}$	(identity 3×3 mtx) $\bar{\mathbf{B}} = \mathbf{A}\mathbf{B} \;\Rightarrow\; \mathbf{A}$ $\mathbf{B}$ is measured
uniform albedo	$\lambda\mathbf{R}$	(orthogonal 3×3 mtx) 6 points: [Drew92] $\lVert\mathbf{A}\mathbf{b}_i\rVert = 1 \Rightarrow \mathbf{b}_i^\top \mathbf{A}^\top \mathbf{A}\mathbf{b}_i = 1 \Rightarrow \mathbf{A}^\top\mathbf{A} \Rightarrow \mathbf{A}$ up to rot. (Choleski)
equal light intensity	$\lambda\mathbf{R}$	$\lVert\mathbf{s}_j\mathbf{A}^{-1}\rVert = 1 \Rightarrow \mathbf{A}$ up to rot. [Hayakawa94]
integrable normals $p_v = q_u$ for $\mathbf{n} \sim (p, q, 1)$	$\begin{bmatrix}\lambda & 0 & \mu \\ 0 & \lambda & \nu \\ 0 & 0 & \tau\end{bmatrix}$	generalized bas-relief ambiguity [Yuille99, Fan97, Belhumeur99]
uniform albedo and integrability	$\lambda\mathbf{I}$	
integrability and 2+ specular pts	$\lambda\mathbf{I}$	[Drbohlav & Chantler, ICCV 2005]

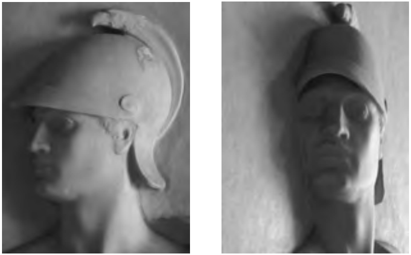
Generalized Bas Relief Ambiguity (GBR)

GBR maps surface $z'(u,v) = \lambda z(u,v) + \mu\, u + \nu\, v$, i.e. it maps normals to $\mathbf{n}' = \mathbf{G}\mathbf{n}$, where

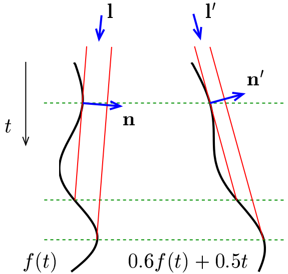
$$\mathbf{G} = \begin{bmatrix} \lambda & 0 & -\mu \\ 0 & \lambda & -\nu \\ 0 & 0 & 1 \end{bmatrix}$$

Obs: If normals change $\mathbf{n}' = \mathbf{G}\mathbf{n}$ and lights change $\mathbf{l}' = \mathbf{G}^{-\top}\,\mathbf{l}$ then Lambertian shading does not change:

$$\mathbf{n}'^{\top}\mathbf{l}' = (\mathbf{n}^{\top}\mathbf{G}^{\top})(\mathbf{G}^{-\top}\mathbf{l}) = \mathbf{n}^{\top}\mathbf{l}$$



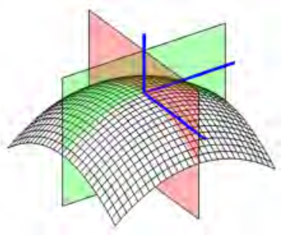
Reproduced from [Belhumeur et al. 1997]



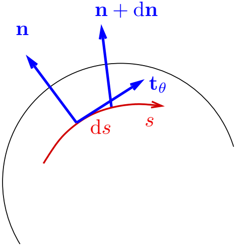
Obs: Shadow boundaries of surface $\mathcal{S}$ illuminated by light $\mathbf{l}$ are identical to those of surface $\mathcal{S}'$ transformed by GBR $\mathbf{G}$ and illuminated by light $\mathbf{l}' = \mathbf{G}^{-\top}\mathbf{l}$

weak assumptions [Belhumeur et al. 1997]

A Quick Glance at the Classical Differential Geometry of Surfaces



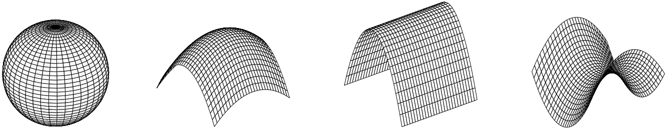
Darboux frame



$\kappa_\theta = \mathbf{t}_\theta^\top \dfrac{\mathrm{d}\mathbf{n}}{\mathrm{d}s}$	normal curvature, direction θ
κ_1, κ_2	principal curvatures
$K = \kappa_1 \cdot \kappa_2$	Gaussian curvature
$H = \kappa_1 + \kappa_2$	mean curvature

$$\kappa_\theta = \kappa_1 \cos^2\theta + \kappa_2 \sin^2\theta$$

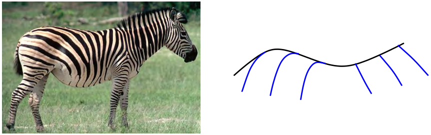
	umbilical	elliptical	parabolic	hyperbolic
convex	$\kappa_1 = \kappa_2 > 0$	$\kappa_1 > 0,\ \kappa_2 > 0$	$\kappa_1 > 0,\ \kappa_2 = 0$	$\kappa_1 > 0,\ \kappa_2 < 0$
concave		$\kappa_1 < 0,\ \kappa_2 < 0$		



the transition elliptic → parabolic → hyperbolic occurs at parabolic lines

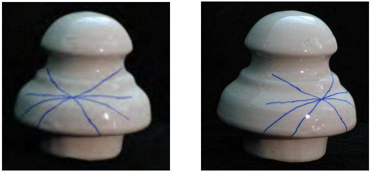
non-umbilical surface like a torus

Occluding Contour Structure

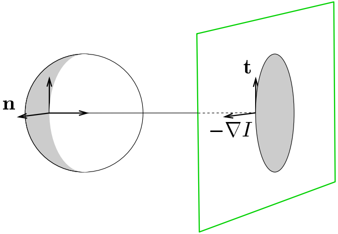


smooth self-occlusion contour (back)
not smooth contour (mane)

- surface curves are tangent to smooth self-occlusion contour



- isophotes are surface curves ⇒ their density approaches infinity on smooth self-occlusion contour



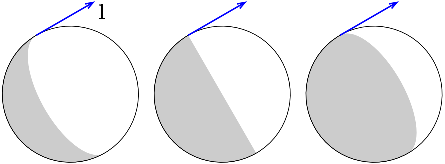
$\mathbf{n} = \mathbf{Q}^\top \underline{\mathbf{t}}$ optical plane normal

$K = \kappa_s \, \kappa_t \quad \rightarrow \quad \operatorname{sign}(K) = \operatorname{sign}(\kappa_t)$

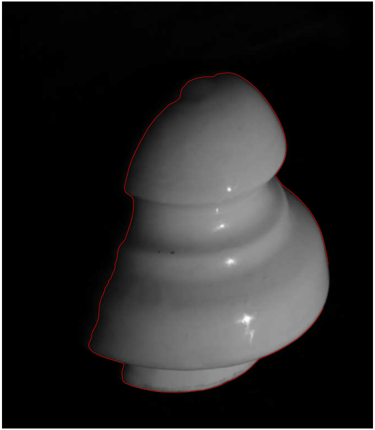
$\kappa_s > 0$ – curvature in the direction of sight
κ_t – occluding contour curvature
$\mathbf{x}_{st} = 0$ since $\mathbf{x}_s \simeq \mathbf{v}$ [Koenderink 84]

- this is a basis for <u>shape from occluding contour</u>

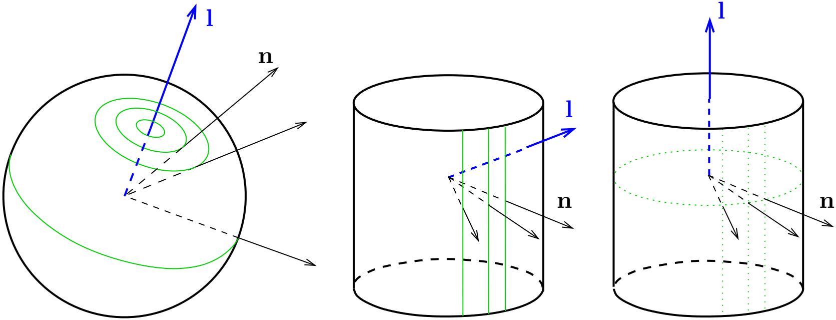
Self-Shadow Contour Structure



- loci where occluding and self-shadow meet: the projection of light direction vector to image plane is tangent to the contour there



Isophotes on Simple Lambertian Surfaces

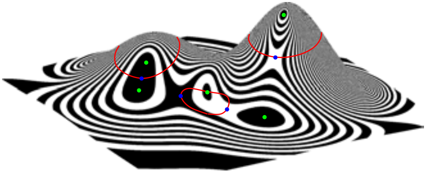


Surface is parameterized by: σ – slant, τ – tilt, where $\mathbf{n}^\top \mathbf{l} = \cos\sigma$

- isophotes – green
- apex – where $\mathbf{n} \simeq \mathbf{l}$
- isophotes parallel to rulings on developable surfaces
- illuminant on cylinder axis: constant reflectance cylindrical part illumination w/o shading
- in general: isophotes are parallel to zero-curvature principal direction

Isophotes on a Complex Surface

shaded Lambertian surface



isophotes w/ approximate parabolic curves

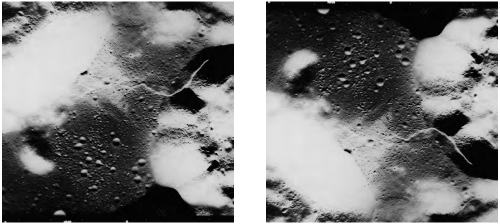
singular image points

- Lambertian apex: move with light, $\mathbf{n} = \mathbf{l}$ (T1)
- extrema and saddles on parabolic lines: move along parabolic lines (T2)
- planar points: do not move (not shown)
- specular points: move with light and/or viewer but slower (not shown)

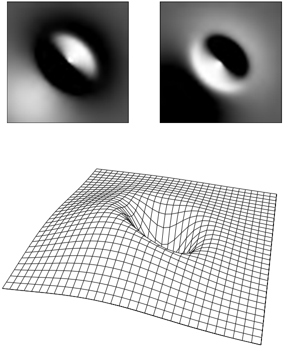
[Koenderink & van Doorn 1980]

The Crater Illusion

Ambiguity in Local Shading and The Human Vision Preference



Apollo 17 landing site (Taurus-Littrow); courtesy of NASA



bottom: crater-like surface
top: surface illuminated from lower-left and top-right

Shading at Lambertian apex:

$$K^2 = \det\left(\mathbf{H}\mathbf{G}^{-1}\right)$$

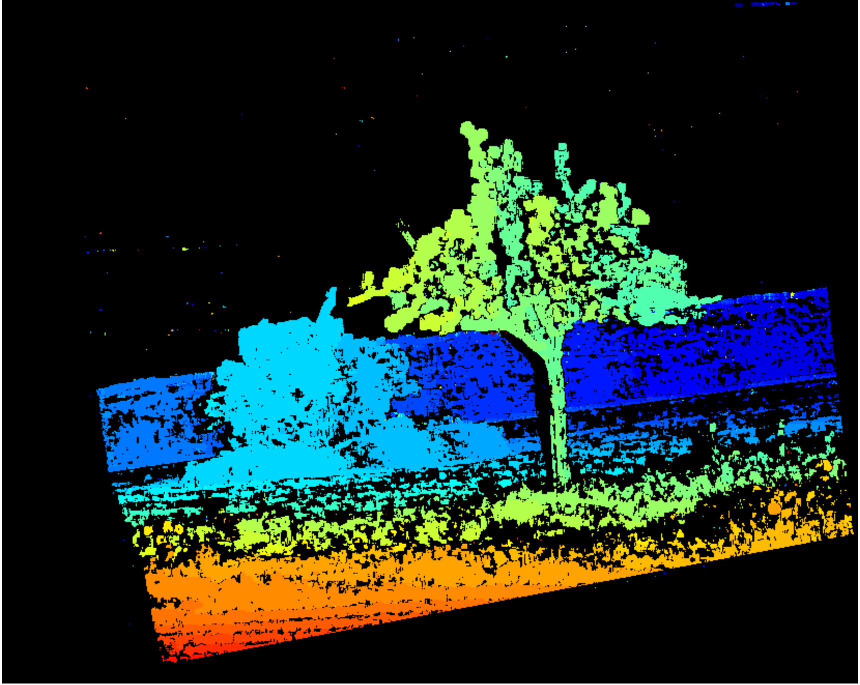
$$2H^2 - K = -\frac{1}{2}\,\mathrm{tr}\left(\mathbf{H}\mathbf{G}^{-1}\right)$$

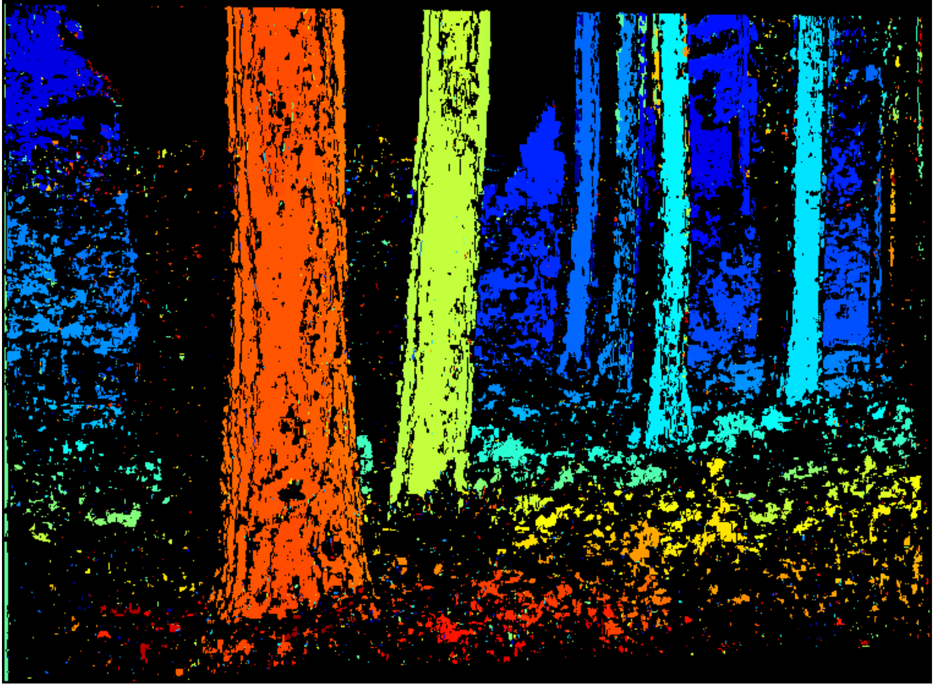
$$\mathbf{H} = \begin{bmatrix} I_{uu} & I_{uv} \\ I_{uv} & I_{vv} \end{bmatrix}$$ image Hessian

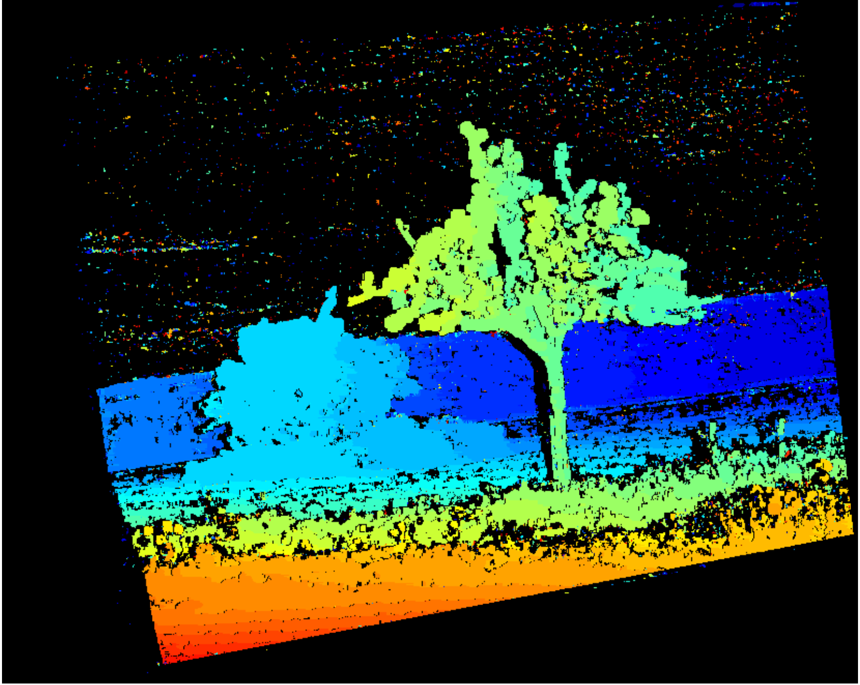
$$\mathbf{G} = \begin{bmatrix} 1 + l_1^2 & l_1 l_2 \\ l_1 l_2 & 1 + l_2^2 \end{bmatrix}$$ from light dir. $\mathbf{l} = (l_1, l_2, l_3)$

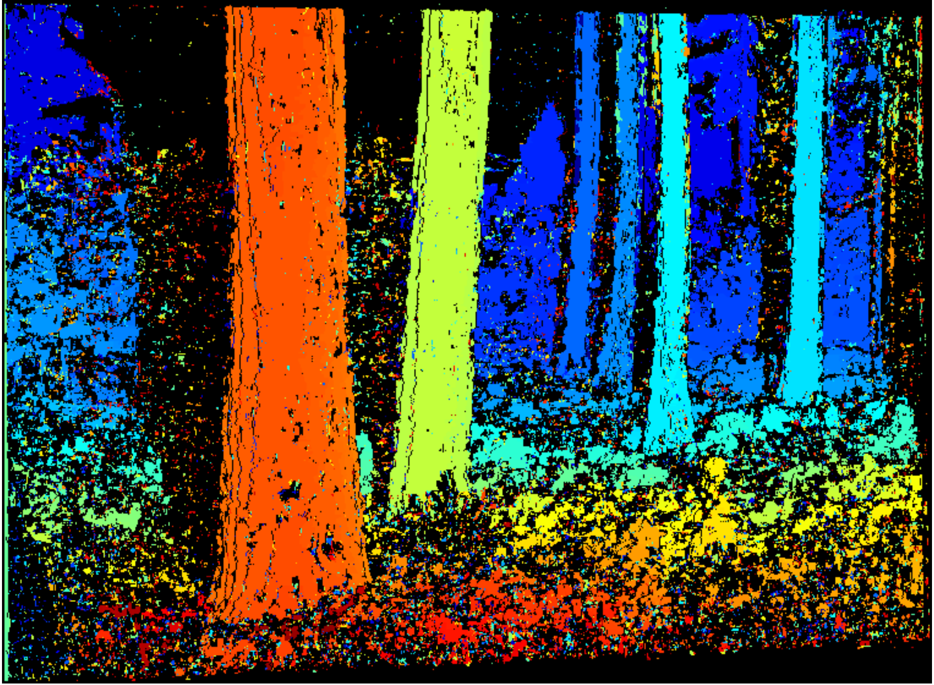
Apex: Up to 4 solutions for surface principal curvatures:
convex/concave × elliptic/hyperbolic

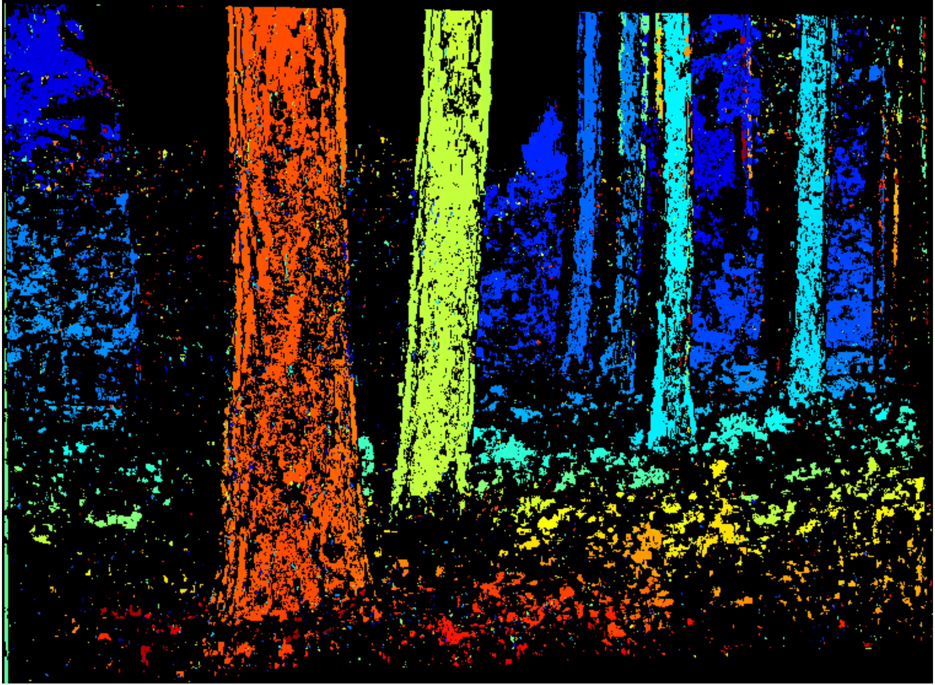
Thank You

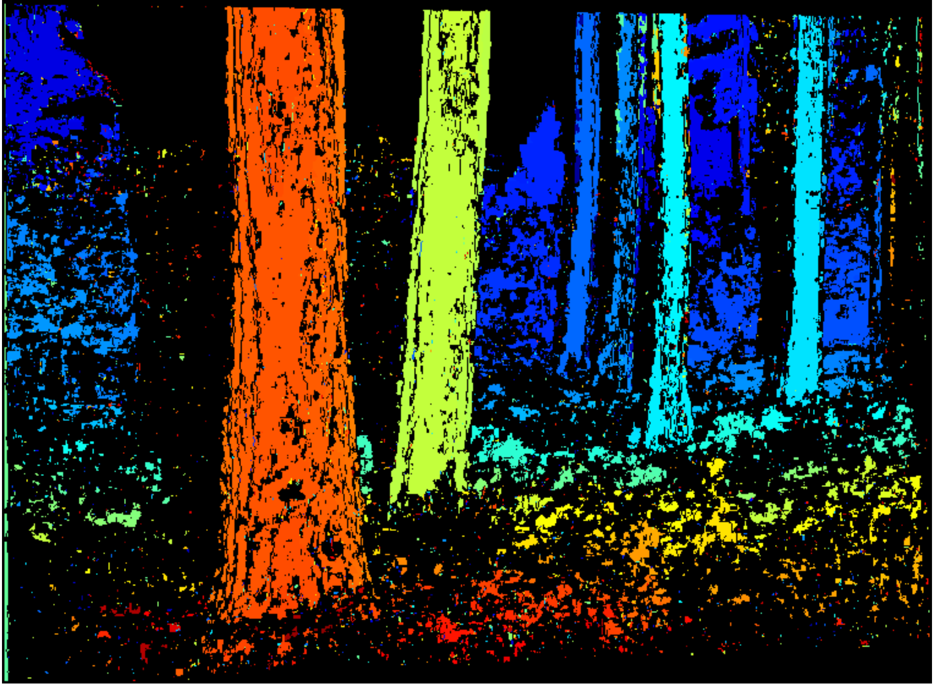


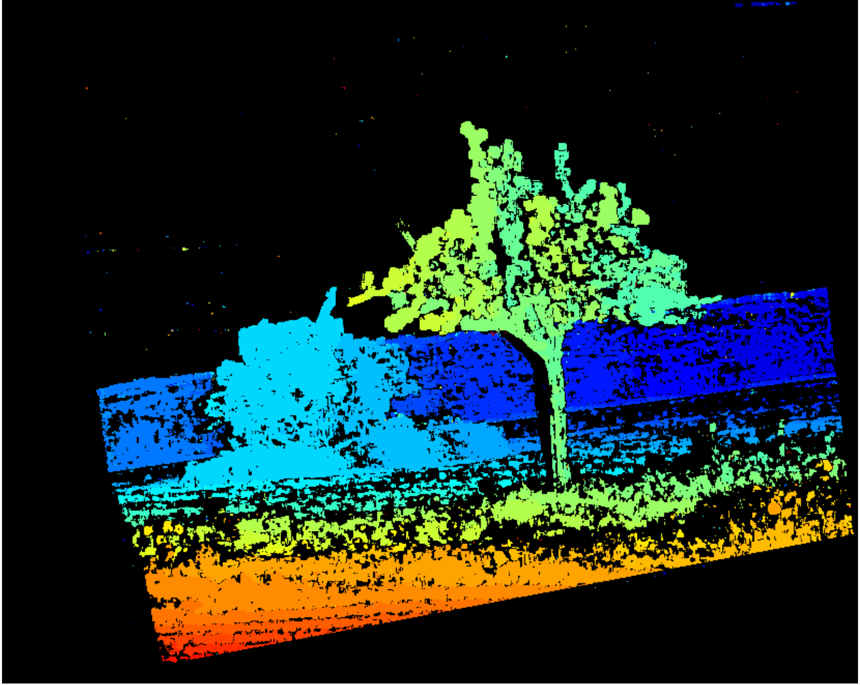


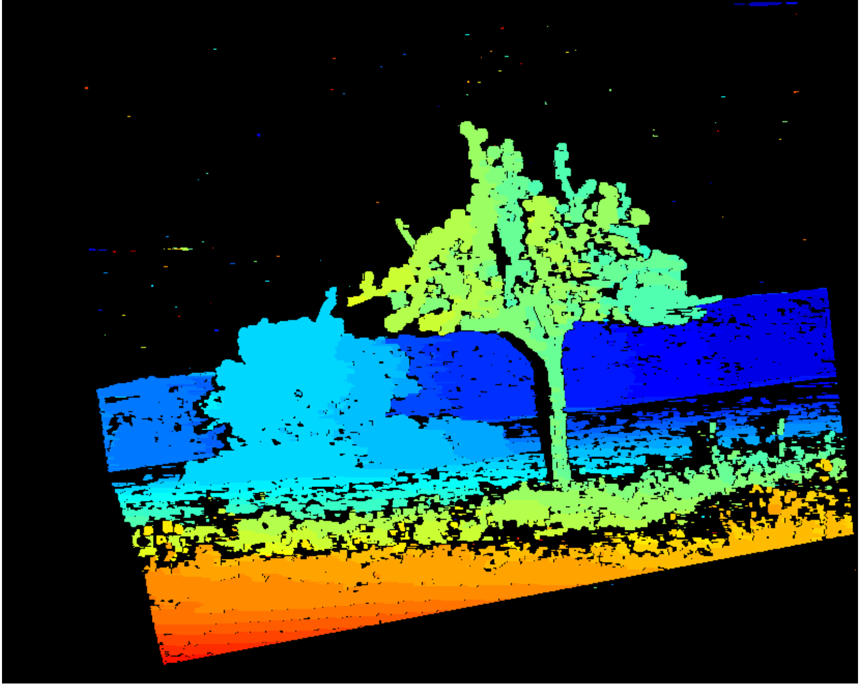


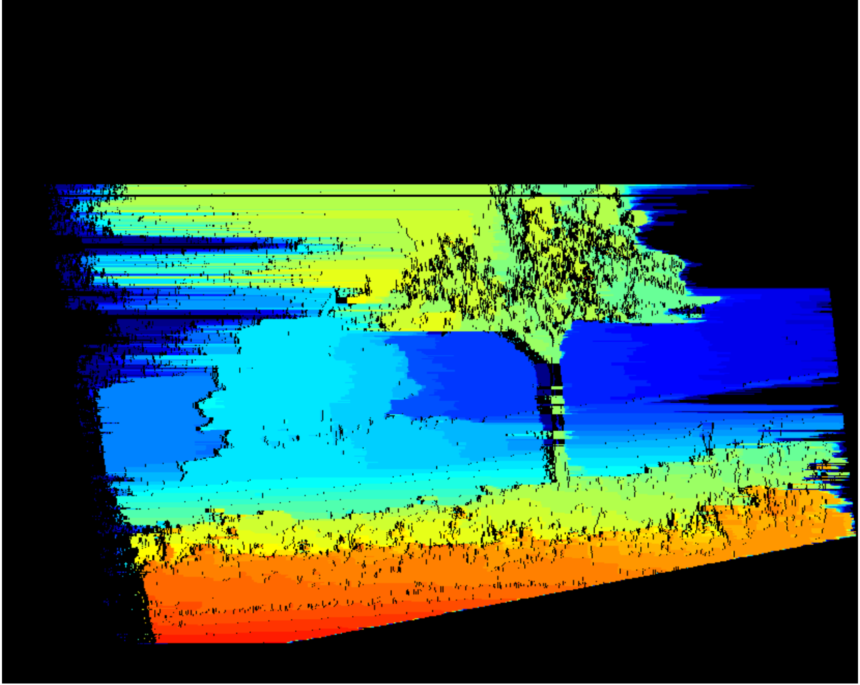


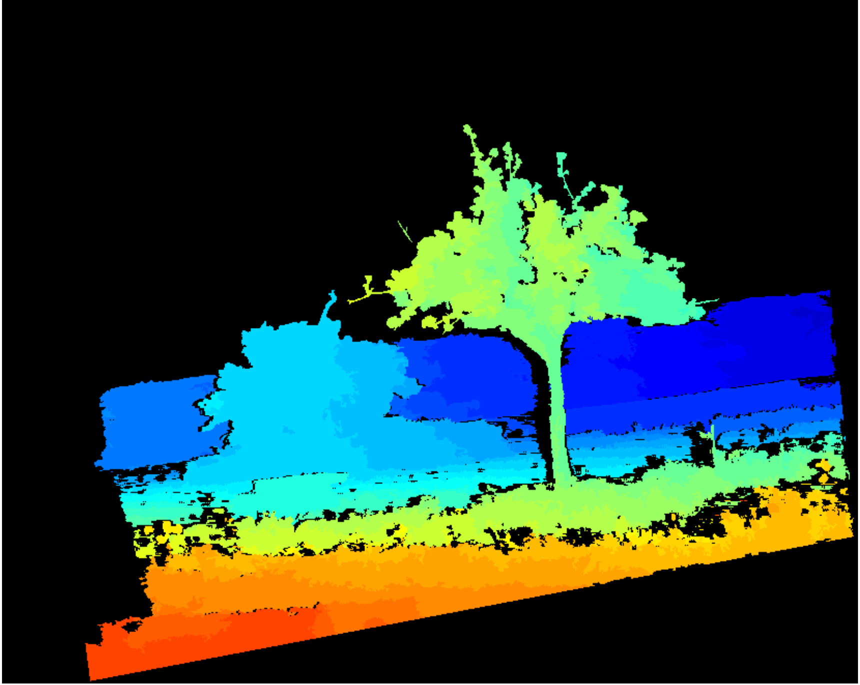


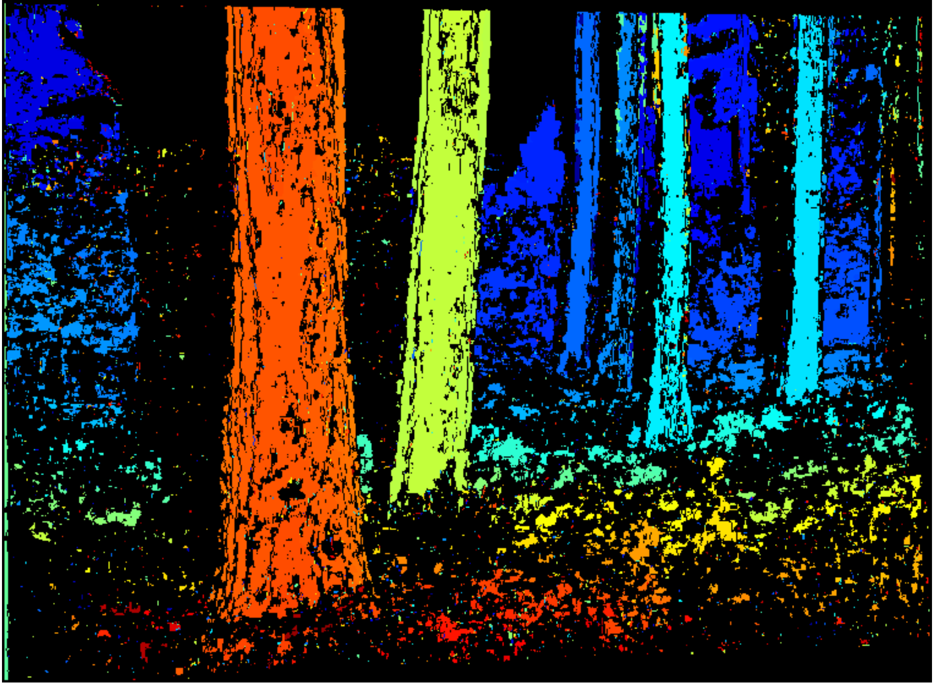


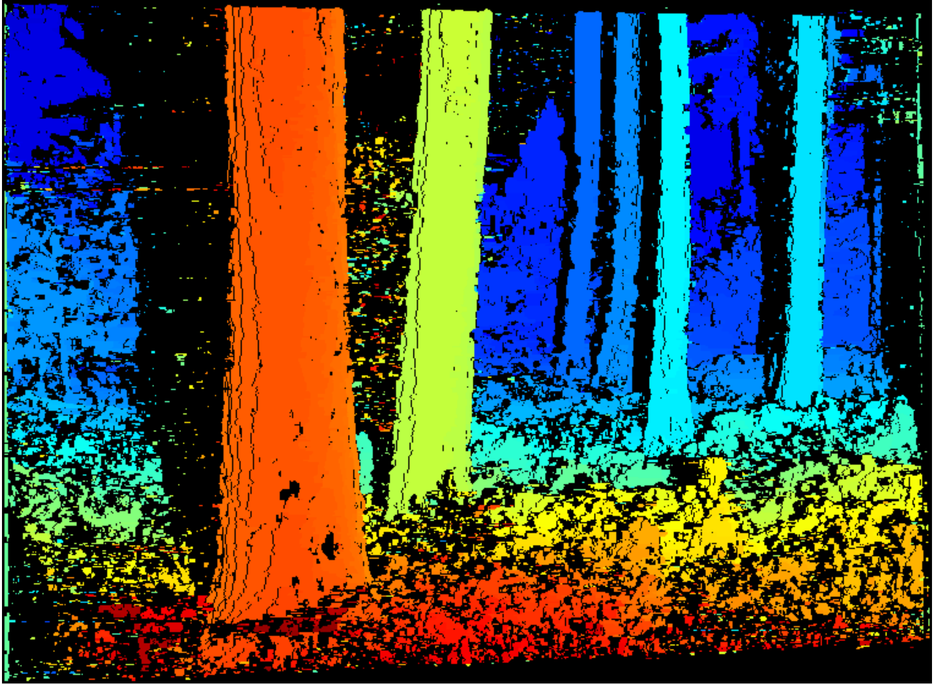


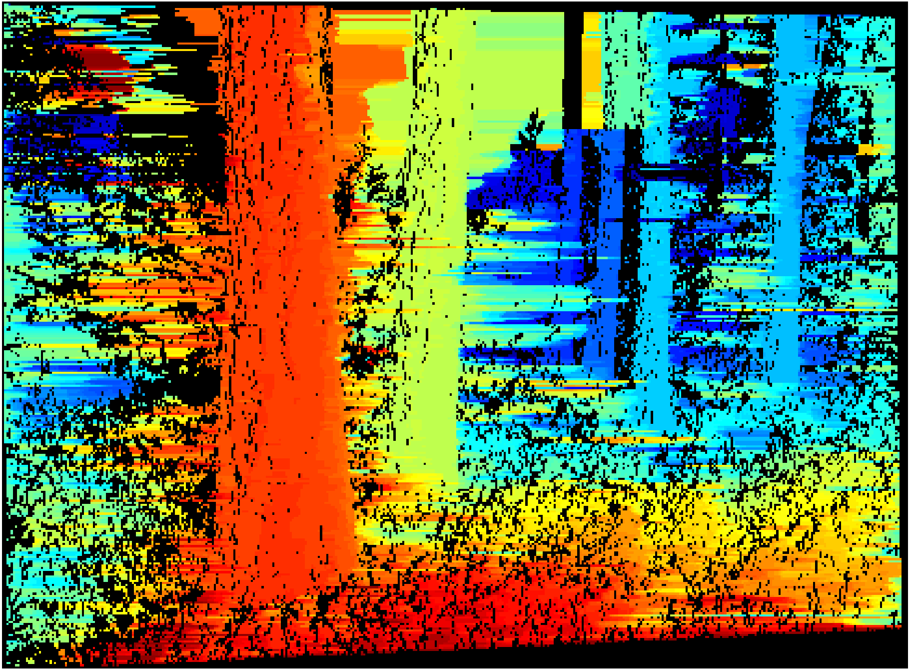


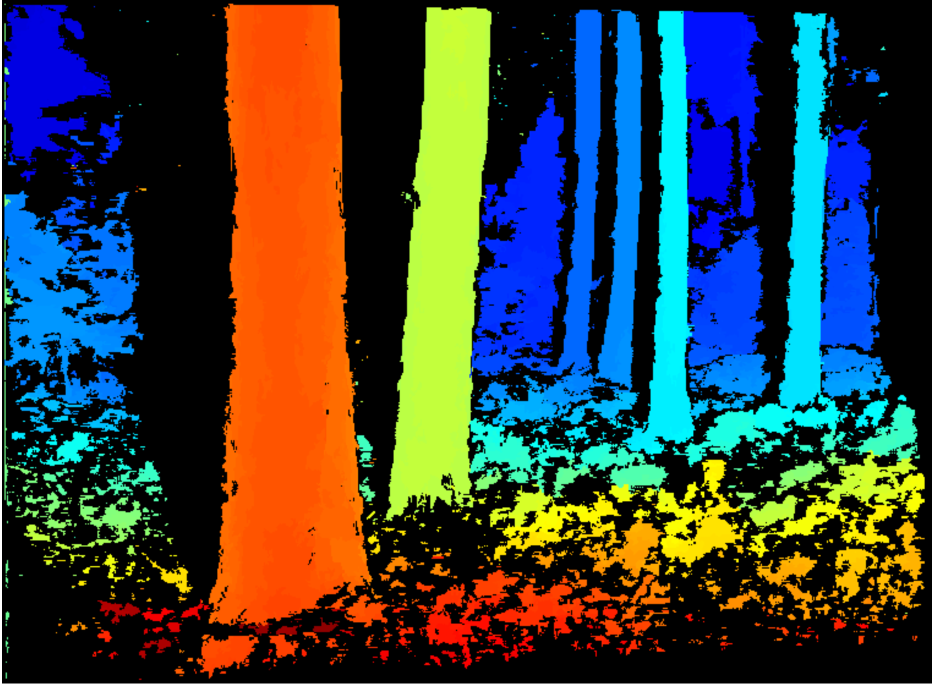




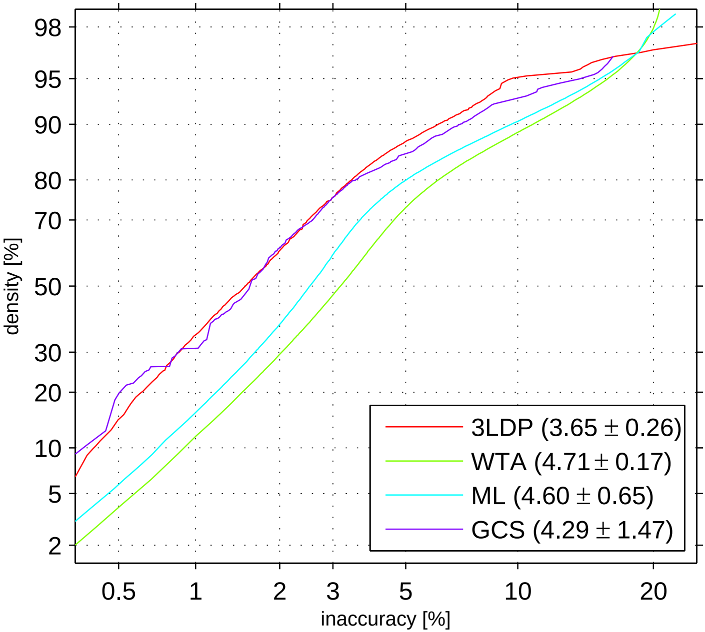


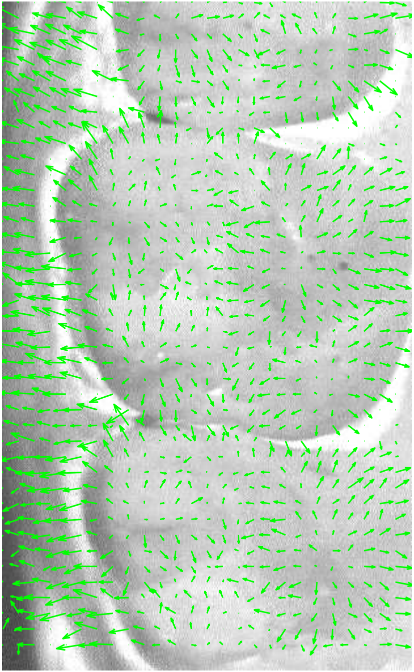


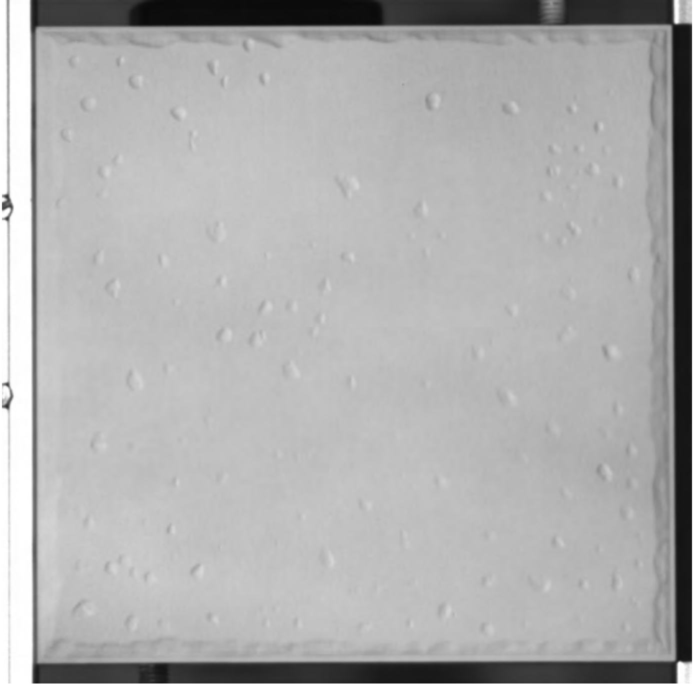


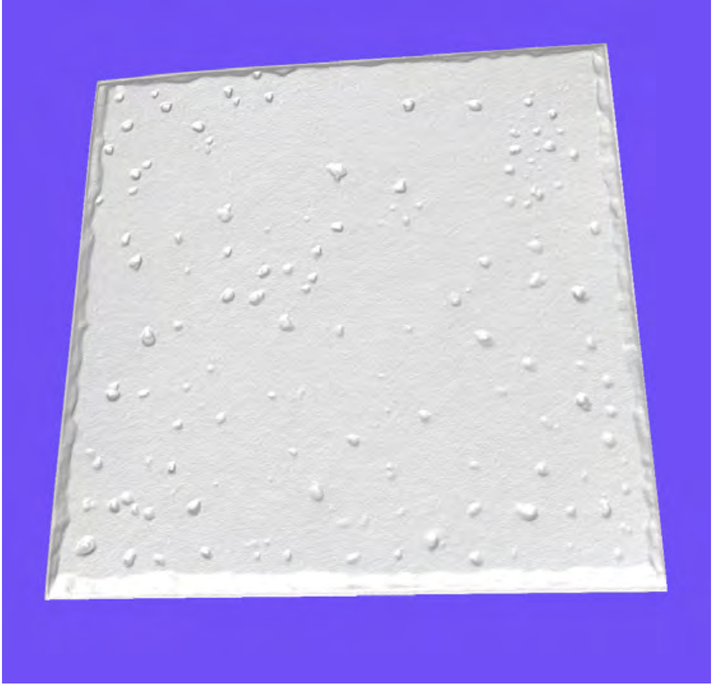


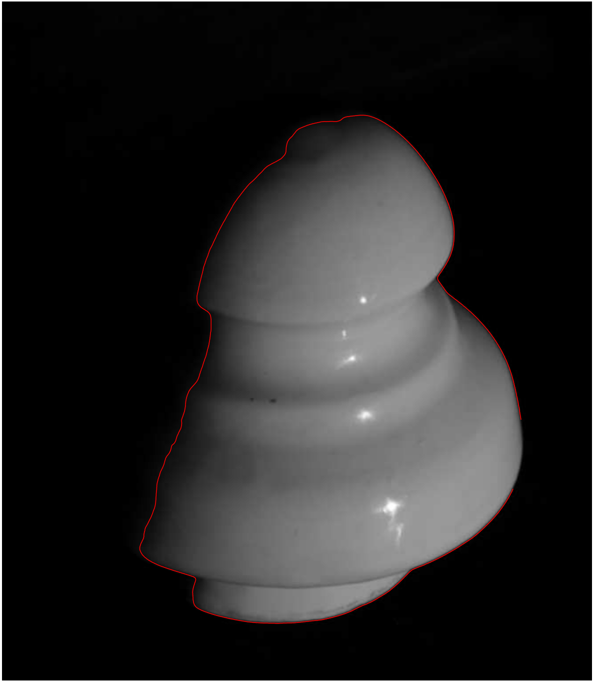
ROC curves and their average error rate bounds

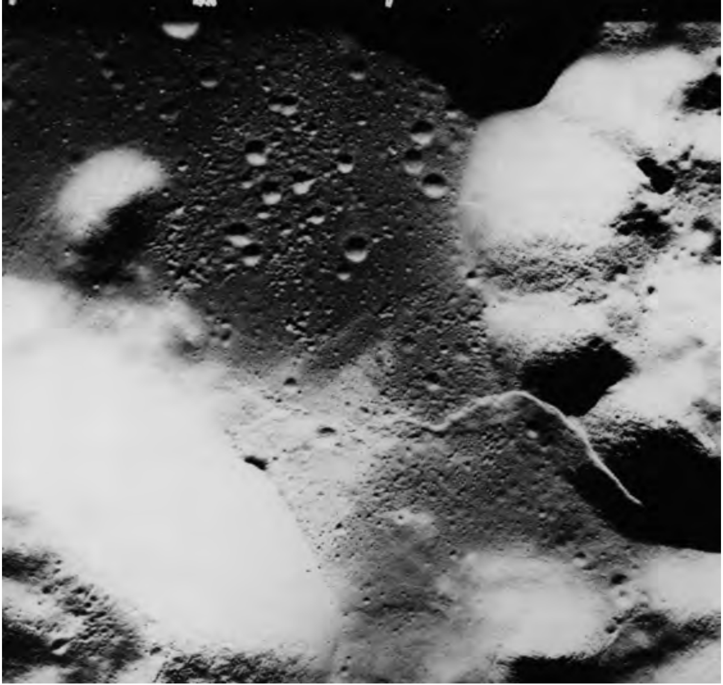












OPPA

PRAHA
PRAGUE
PRAGA
PRAG

EVROPSKÁ UNIE

OPPA European Social Fund
Prague & EU: We invest in your future.