TUTORIAL

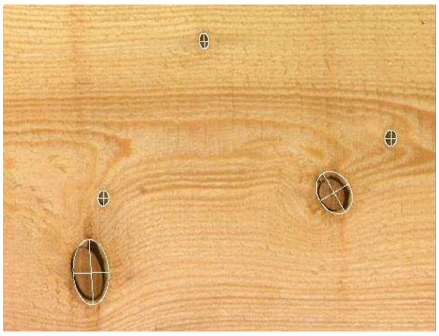
CVPR 2011
June 24, 2011

Image and Video Description with Local Binary Pattern Variants

Prof. Matti Pietikäinen, Prof. Janne Heikkilä
{mkp,jth}@ee.oulu.fi

Machine Vision Group
University of Oulu, Finland
http://www.cse.oulu.fi/MVG/

Texture is an important characteristic of images and videos



Contents

Part 1: Introduction to local binary patterns in spatial and spatiotemporal domains

Matti Pietikäinen

LBP in spatial domain

2-D surface texture is a two dimensional phenomenon characterized by:

- spatial structure (pattern)
- contrast ('amount' of texture)

Transformation	Property: Pattern	Property: Contrast
Gray scale	no effect	affects
Rotation	affects	no effect
Zoom in/out	affects	?

Thus,
1) contrast is of no interest in gray scale invariant analysis
2) often we need a gray scale and rotation invariant pattern measure

Local Binary Pattern and Contrast operators

Ojala T, Pietikäinen M & Harwood D (1996) A comparative study of texture measures with classification based on feature distributions. Pattern Recognition 29:51-59.

An example of computing LBP and C in a 3x3 neighborhood:

example

6	5	2
7	6	1
9	8	7

thresholded

1	0	0
1		0
1	1	1

weights

1	2	4
128		8
64	32	16

Pattern = **11110001**

LBP = 1 + 16 +32 + 64 + 128 = **241**

C = (6+7+8+9+7)/5 - (5+2+1)/3 = **4.7**

Important properties:

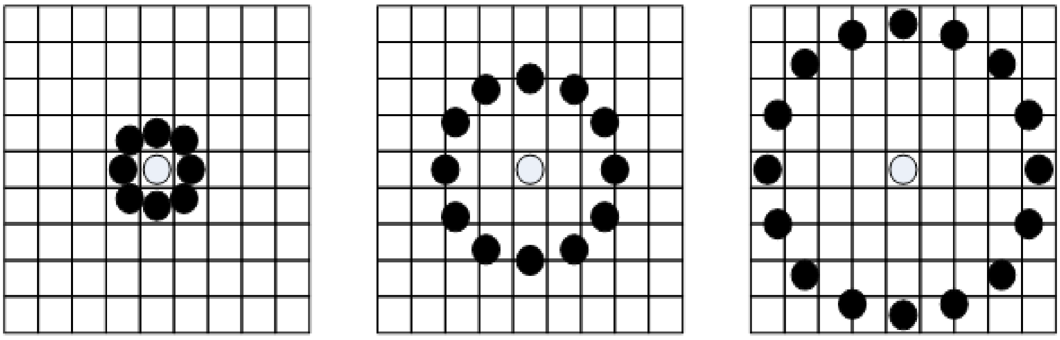
- LBP is invariant to any monotonic gray level change
- computational simplicity

Infotech Oulu
MACHINE VISION GROUP
UNIVERSITY of OULU
OULUN YLIOPISTO

Multiscale LBP

Ojala T, Pietikäinen M & Mäenpää T (2002) Multiresolution gray-scale and rotation invariant texture classification with Local Binary Patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence 24(7):971-987.

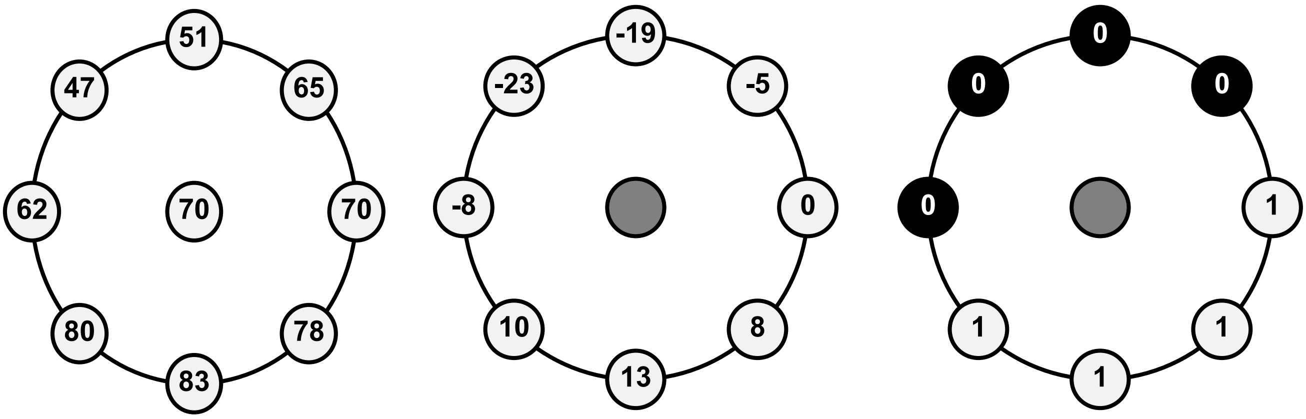
- arbitrary circular neighborhoods
- uniform patterns
- multiple scales
- rotation invariance
- gray scale variance as contrast measure



P = 8, R = 1.0 | P = 12, R = 2.5 | P = 16, R = 4.0

The value of the LBP code of a pixel (x_c, y_c) is given by:

$$LBP_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c)2^p \qquad s(x) = \begin{cases} 1, if\ x \geq 0; \\ 0, otherwise. \end{cases}$$

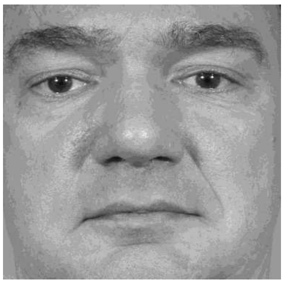


1. Sample | 2. Difference | 3. Threshold

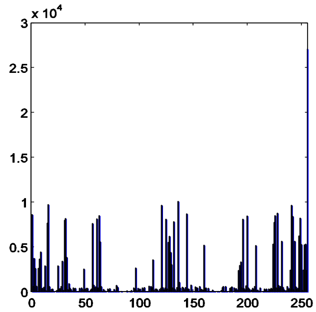
$\boxed{1*1 + 1*2 + 1*4 + 1*8 + 0*16 + 0*32 + 0*64 + 0*128 = 15}$

4. Multiply by powers of two and sum

An example of LBP image and histogram


Input image

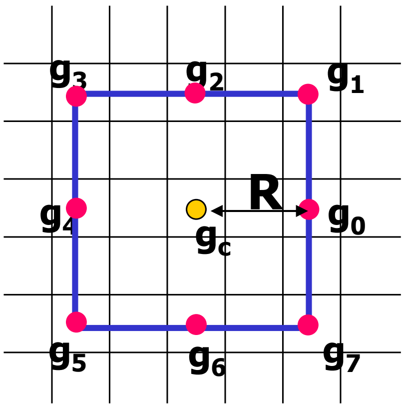

LBP image



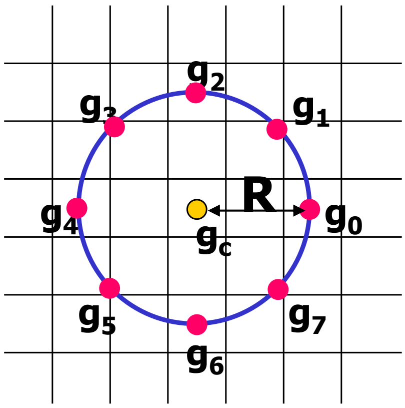
LBP histogram

Foundations for LBP: Description of local image texture

Texture at g_c is modeled using a local neighborhood of radius R, which is sampled at P (8 in the example) points:



Square neighborhood



Circular neighborhood (g_1,g_3,g_5,g_7 interpolated)

Let's define texture T as the joint distribution of gray levels g_c and g_p (p=0,…, P-1):

$$T = t(g_c, g_0, \ldots, g_{P-1})$$

Description of local image texture (cont.)

Without losing information, we can subtract g_c from g_p:

$$T = t(g_c, g_0 - g_c, \ldots, g_{P-1} - g_c)$$

Assuming g_c is independent of $g_p - g_c$, we can factorize above:

$$T \sim t(g_c)\, t(g_0 - g_c, \ldots, g_{P-1} - g_c)$$

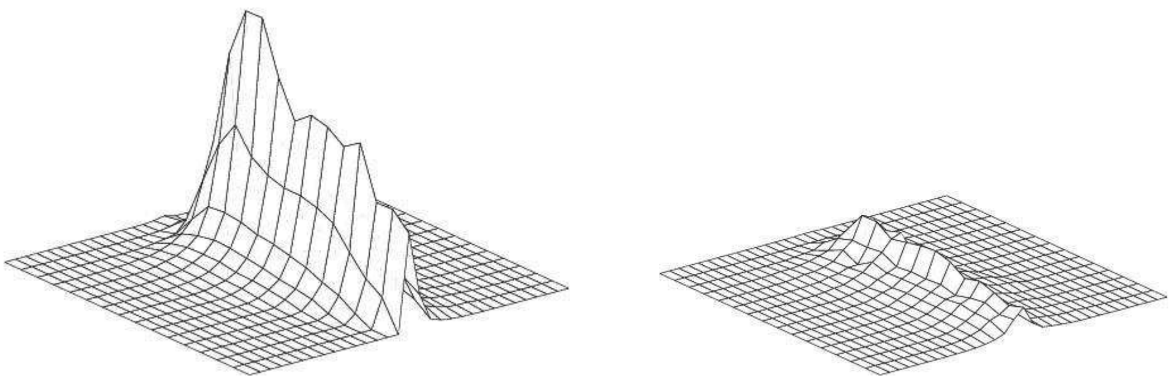
$t(g_c)$ describes the overall luminance of the image, which is unrelated to local image texture, hence we ignore it:

$$T \sim t(g_0 - g_c, \ldots, g_{P-1} - g_c)$$

Above expression is invariant wrt. gray scale shifts

Description of local image texture (cont.)

Exact independence of $t(g_c)$ and $t(g_0 - g_c, \ldots, g_{P-1} - g_c)$ is not warranted in practice:



average $t(g_c, g_0 - g_c)$

average absolute difference between $t(g_c, g_0 - g_c)$ and $t(g_c)\, t(g_0 - g_c)$

Pooled (G=16) from 32 Brodatz textures used in [Ojala, Valkealahti, Oja & Pietikäinen: Pattern Recognition 2001]

LBP: Local Binary Pattern

Invariance wrt. any monotonic transformation of the gray scale is achieved by considering the signs of the differences:

$$T \sim t(s(g_0-g_c), \ldots, s(g_{P-1}-g_c))$$

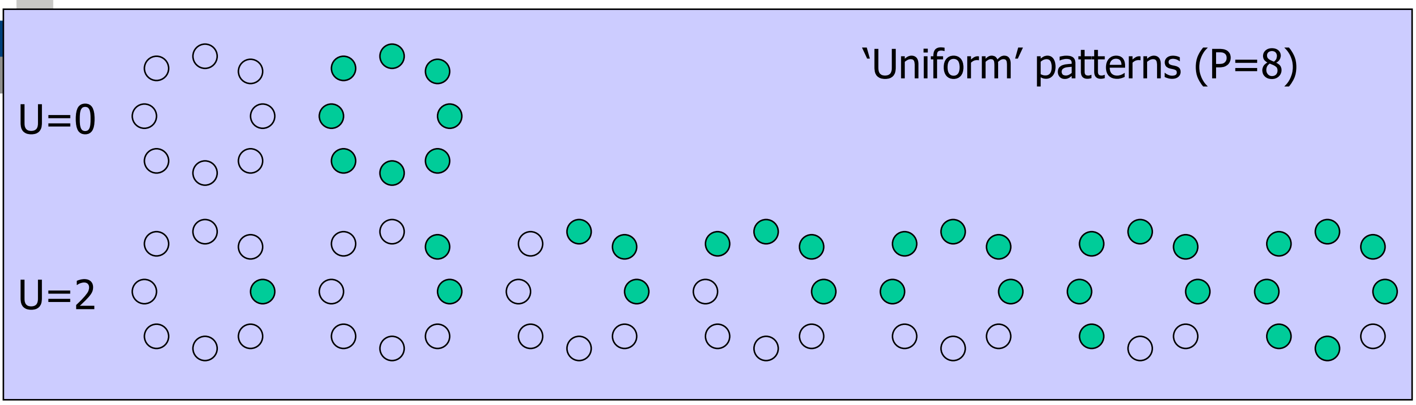
where

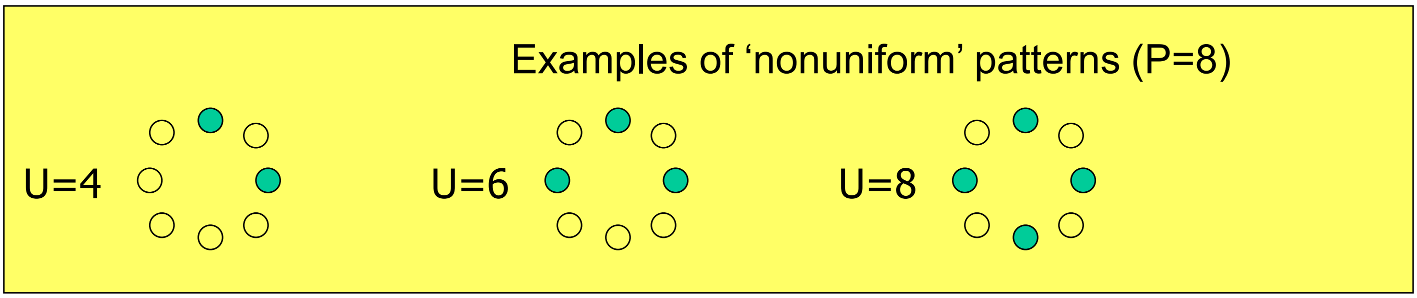
$$s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

Above is transformed into a unique P-bit pattern code by assigning binomial coefficient 2^p to each sign $s(g_p-g_c)$:

$$LBP_{P,R} = \sum_{p=0}^{P-1} s\,(g_p-g_c)\,2^p$$

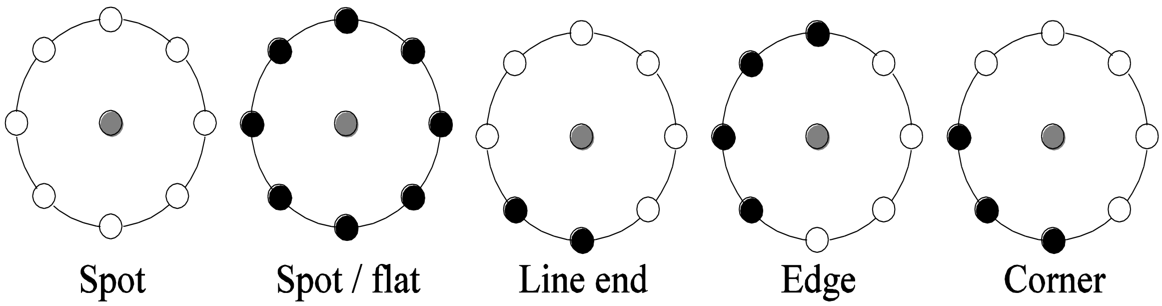
'Uniform' patterns





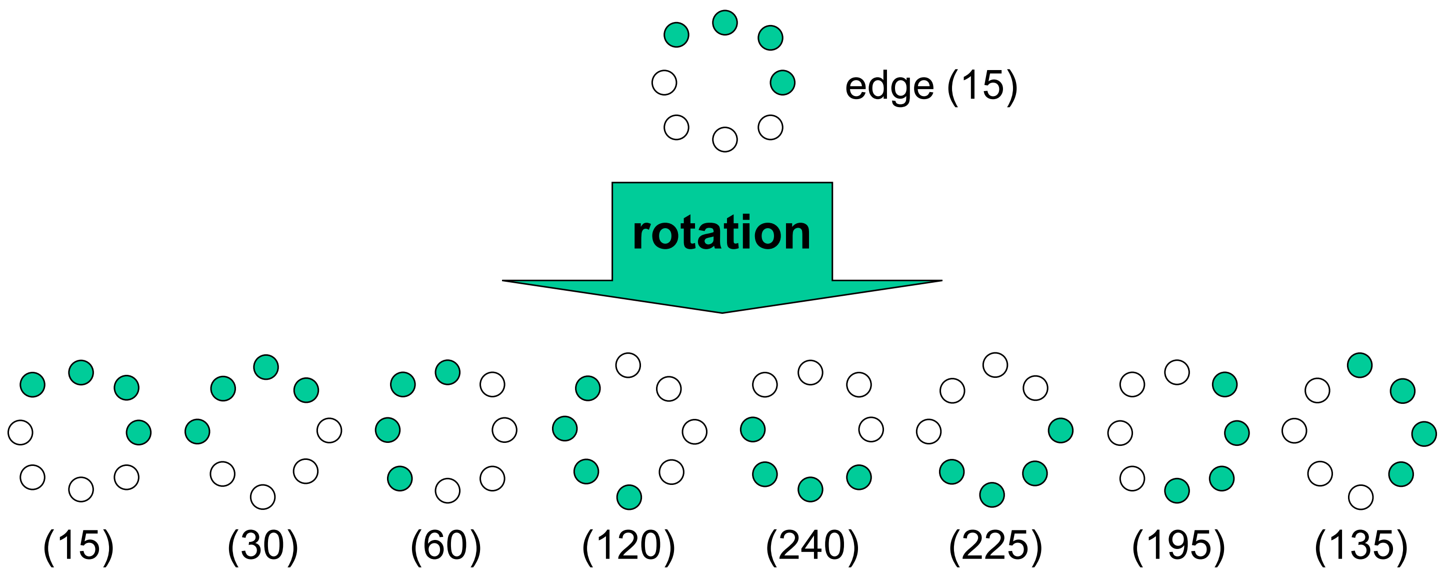
Texture primitives ("micro-textons") detected by the uniform patterns of LBP

1 = black
0 = white



Rotation of Local Binary Patterns

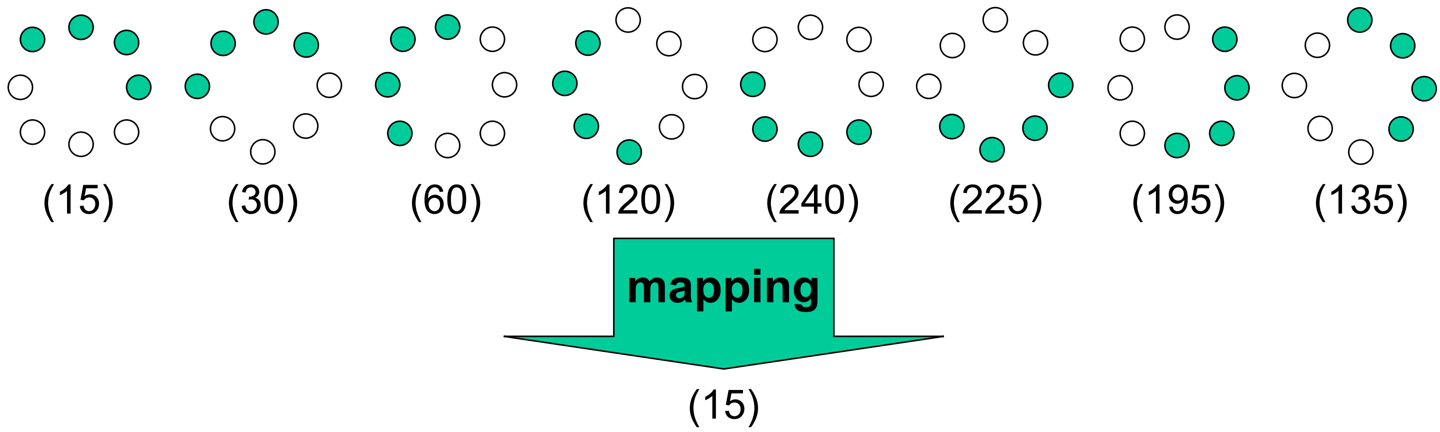
Spatial rotation of the binary pattern changes the LBP code:



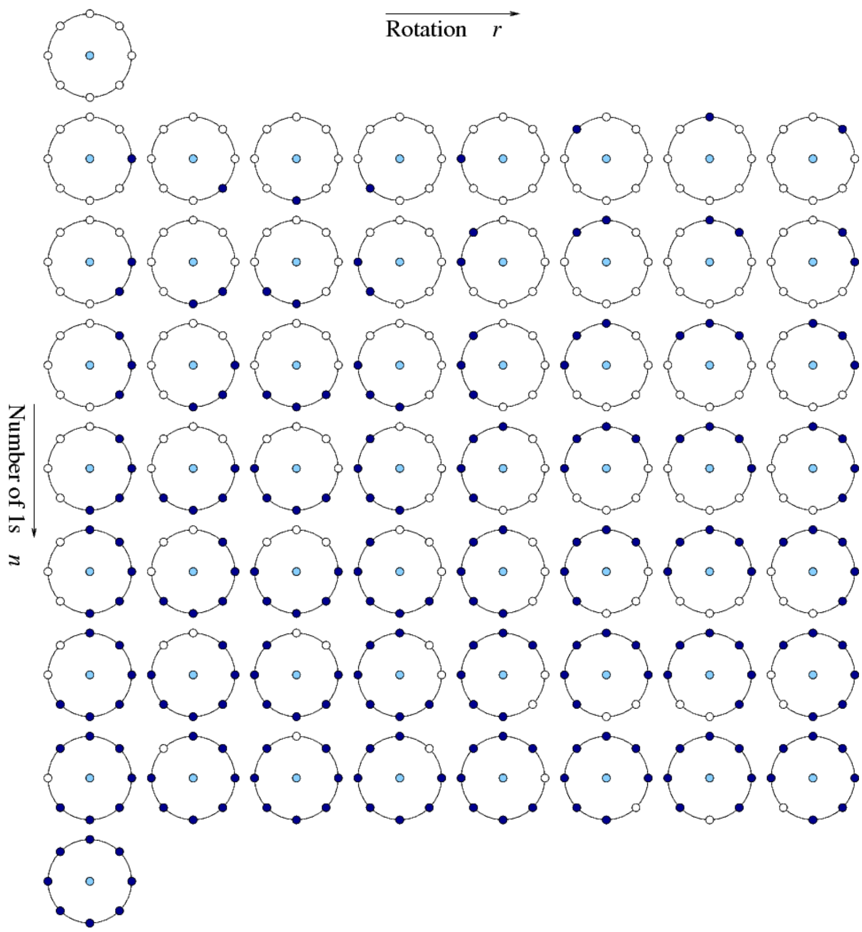
Rotation invariant local binary patterns

Formally, rotation invariance can be achieved by defining:

$$LBP_{P,R}^{ri} = \min\{ROR(LBP_{P,R}, i) \mid i=0, \ldots, P-1\}$$



- Bit patterns with 0 or 2 transitions *0→1* or *1→0* when the pattern is considered circular
- All non-uniform patterns assigned to a single bin
- 58 uniform patterns in case of 8 sampling points



Operators for characterizing texture contrast

Local gray level variance can be used as a contrast measure:

$$VAR_{P,R} = \frac{1}{P} \sum_{p=0}^{P-1} (g_p - m)^2$$

where

$$m = \frac{1}{P} \sum_{p=0}^{P-1} g_p$$

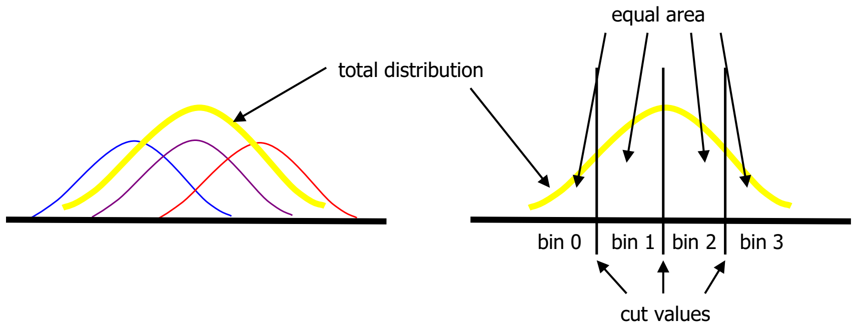
$VAR_{P,R}$
- invariant wrt. gray scale shifts (but not to any monotonic transformation like LBP)
- invariant wrt. rotation along the circular neighborhood

Usually using complementary contrast leads to a better performance than using LBP alone, but this is ignored in many comparative studies!

Quantization of continuous feature space

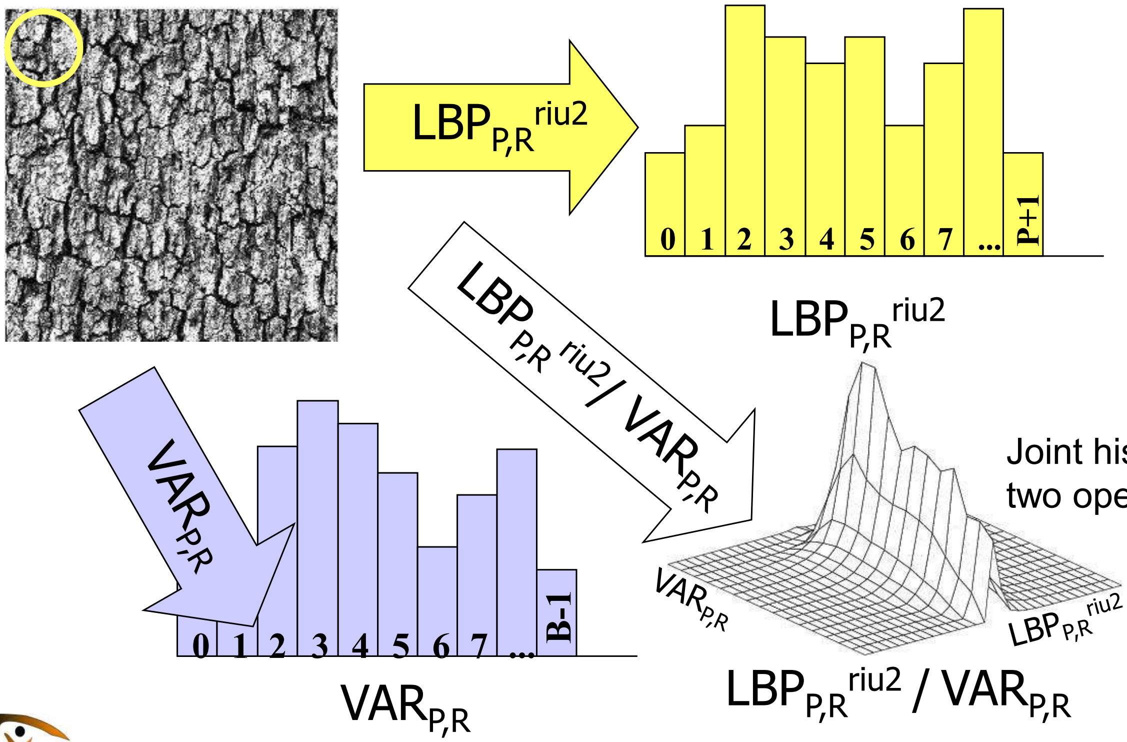
Texture statistics are described with discrete histograms
- Mapping needed for continuous-valued contrast features

Nonuniform quantization
- Every bin have the same amount of total data
- Highest resolution of the quantization is used where the number of entries is largest



Estimation of empirical feature distributions

Input image (region) is scanned with the chosen operator(s), pixel by pixel, and operator outputs are accumulated into a discrete histogram



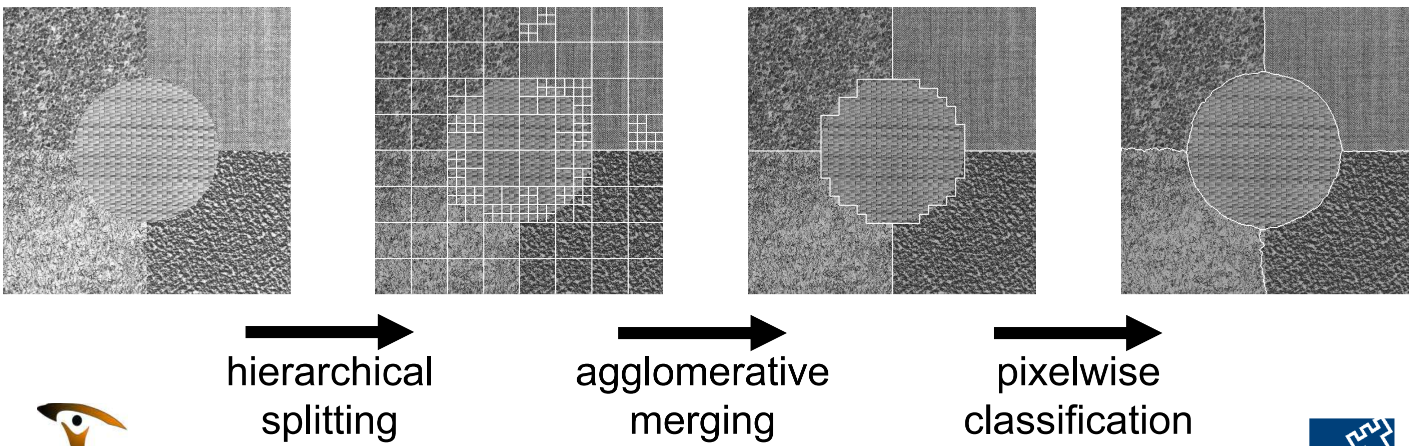
Example problem: Unsupervised texture segmentation

Ojala T & Pietikäinen M (1999) Unsupervised texture segmentation using feature distributions. Pattern Recognition 32:477-486.

- LBP/C was used as texture operator

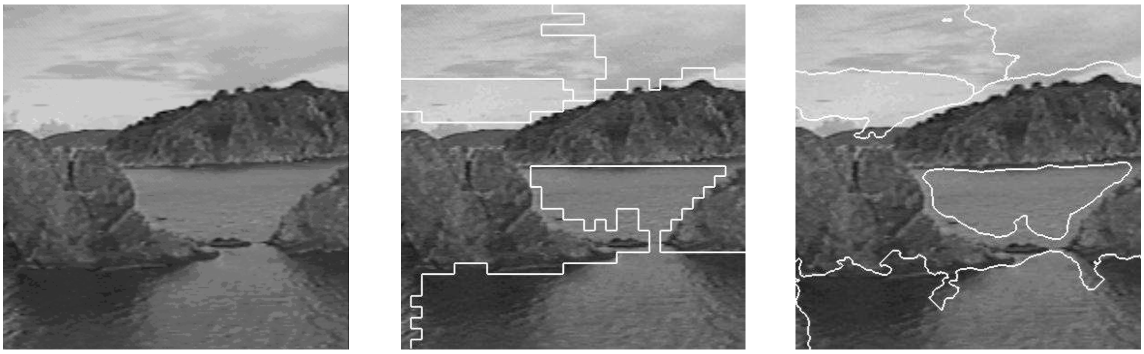
Segmentation algorithm consists of three phases:

1. hierarchical splitting
2. agglomerative merging
3. pixelwise classification

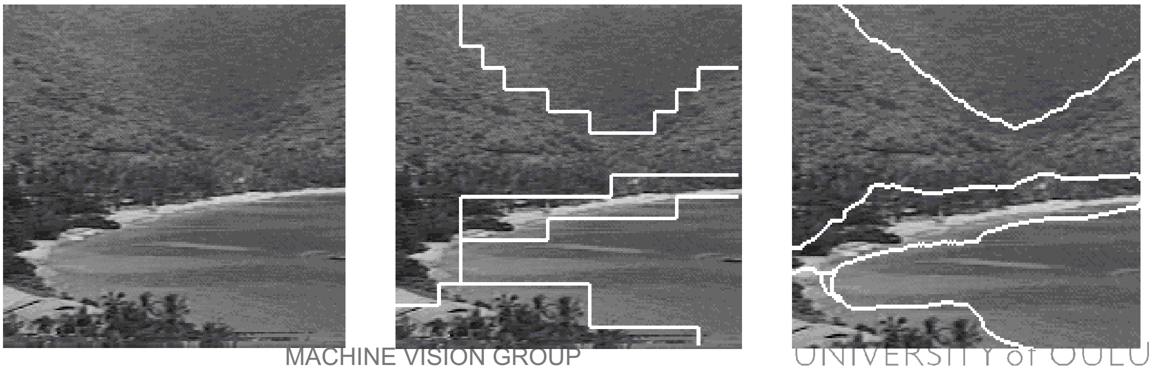


Segmentation examples

Natural scene #1: 384x384 pixels



Natural scene #2: 192x192 pixels



Multiscale analysis

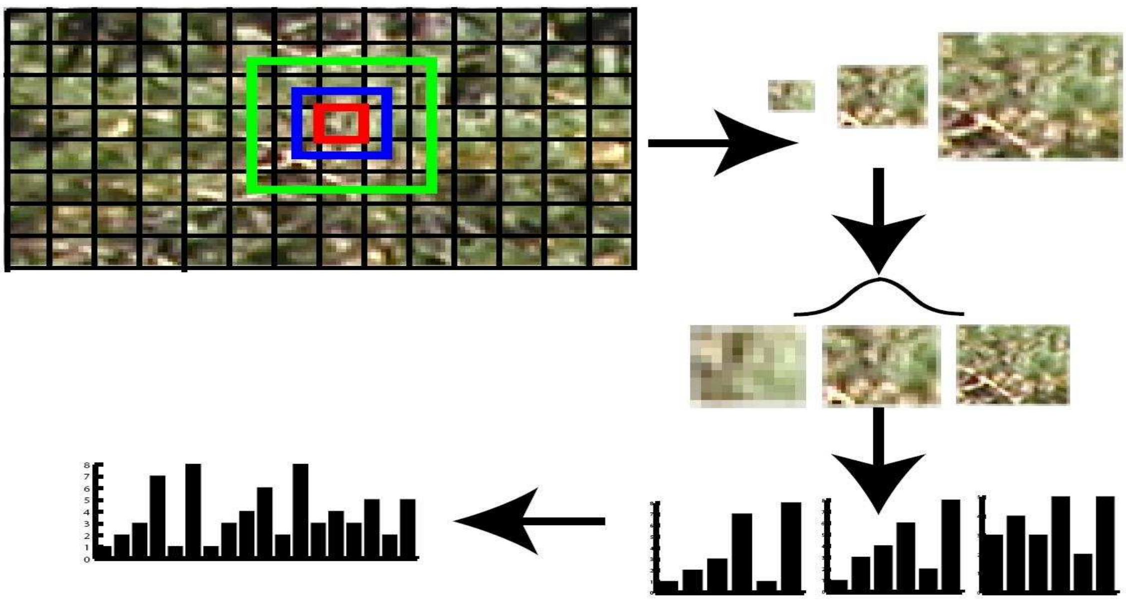
Information provided by N operators can be combined simply by summing up operatorwise similarity scores into an aggregate similarity score:

$$L_N = \sum_{n=1}^{N} L_n \quad \text{e.g.} \quad LBP_{8,1}^{riu2} + LBP_{8,3}^{riu2} + LBP_{8,5}^{riu2}$$

Effectively, the above assumes that distributions of individual operators are independent

Multiscale analysis using images at multiple scales

Image regions can be e.g. re-scaled prior to feature extraction



Nonparametric classification principle

In Nearest Neighbor classification, sample S is assigned to the class of model M that maximizes

$$L(S,M) = \sum_{b=0}^{B-1} S_b \ln M_b$$

Instead of log-likelihood statistic, chi square distance or histogram intersection is often used for comparing feature distributions.

The histograms should be normalized e.g. to unit length before classification, if the sizes of the image windows to be analyzed can vary.

The bins of the LBP feature distribution can also be used directly as features e.g. for SVM classifiers.

Infotech Oulu
MACHINE VISION GROUP
UNIVERSITY of OULU
OULUN YLIOPISTO

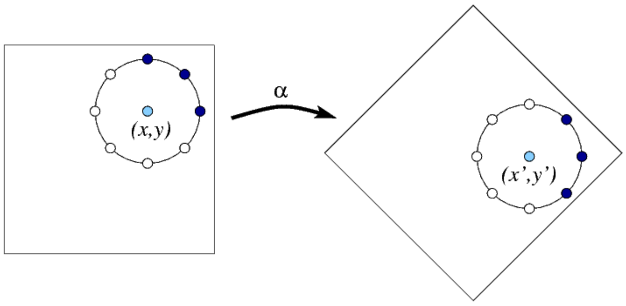
LBP histogram Fourier features

Ahonen T, Matas J, He C & Pietikäinen M (2009) Rotation invariant image description with local binary pattern histogram fourier features. In: Image Analysis, SCIA 2009 Proceedings, Lecture Notes in Computer Science 5575, 61-70.

Rotation revisited

Rotation of an image by α degrees

- Translates each local neighborhood to a new location
- Rotates each neighborhood by α degrees



If α = 45°, local binary patterns

00000001 → 00000010,

00000010 → 00000100, ...,

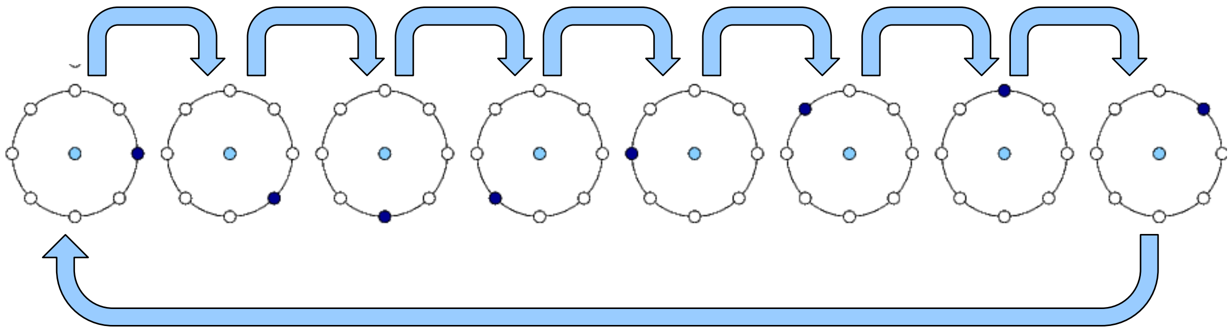
11110000 → 11100001, ...,

Similarly if α = *k**45°,

each pattern is circularly rotated by *k* steps

Rotation revisited (2)

In the uniform LBP histogram, rotation of input image by *k**45° causes a cyclic shift by *k* along each row:



Rotation invariant features

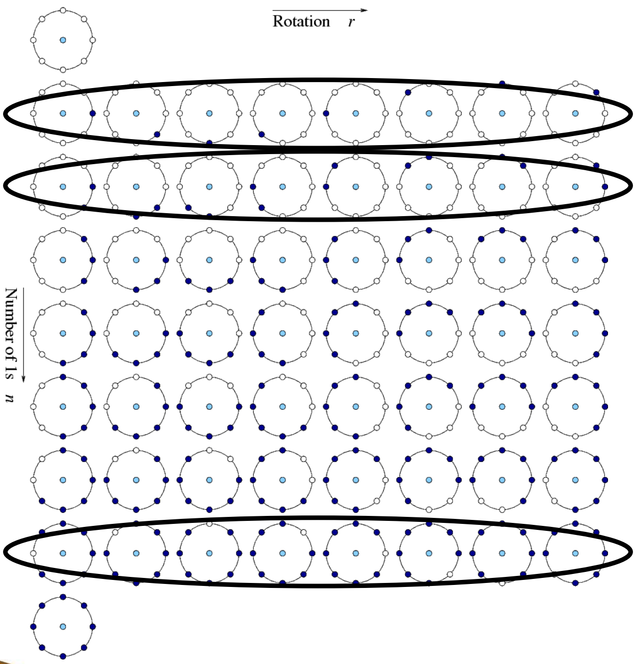
LBP histogram features that are
invariant to cyclic shifts along the rows are
invariant to *k**45° rotations of the input image

- Sum (original rotation invariant LBP)
- Cyclic autocorrelation
- Rapid transform
- **Fourier magnitude**

LBP Histogram Fourier Features

$$H(n,u) = \sum_{r=0}^{P-1} h_I(U_P(n,r)) e^{-i2\pi ur/P}$$

$$|H(n,u)| = \sqrt{H(n,u)\overline{H(n,u)}}$$



Fourier magnitudes

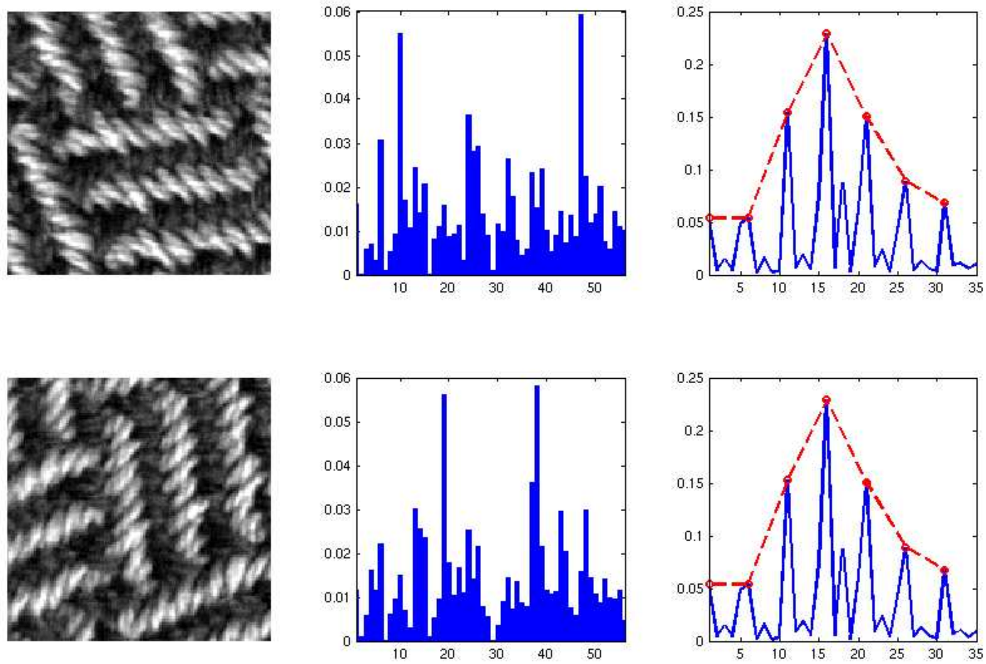
Fourier magnitudes

.
.
.

Fourier magnitudes

LBP-HF feature vector

Example



Input image

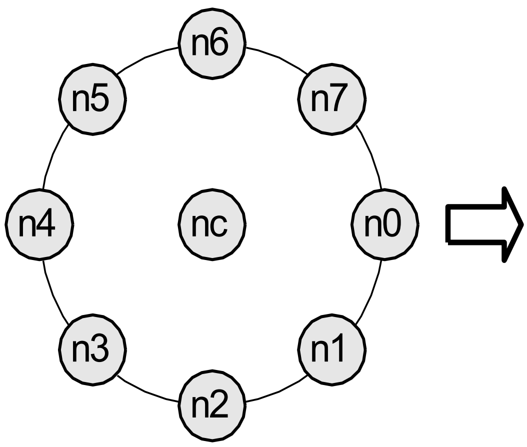
Uniform
LBP histogram

Original rot.invariant LBP (red)
LBP-Histogram fourier (blue)

Description of interest regions with center-symmetric LBPs

Heikkilä M, Pietikäinen M & Schmid C (2009) Description of interest regions with local binary patterns. Pattern Recognition 42(3):425-436.

Neighborhood

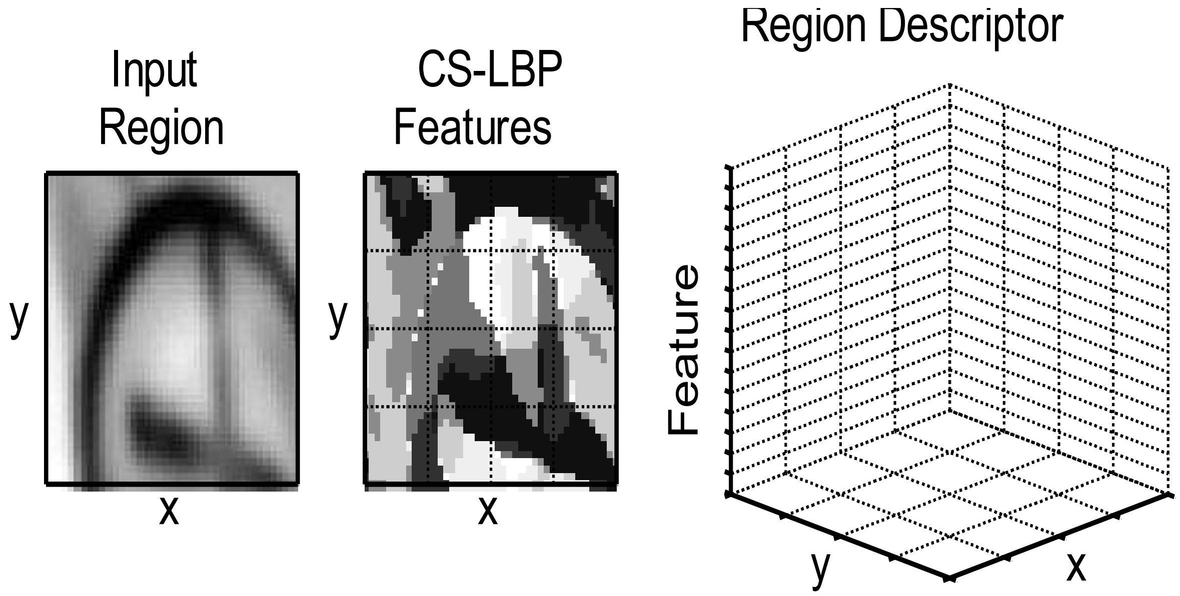


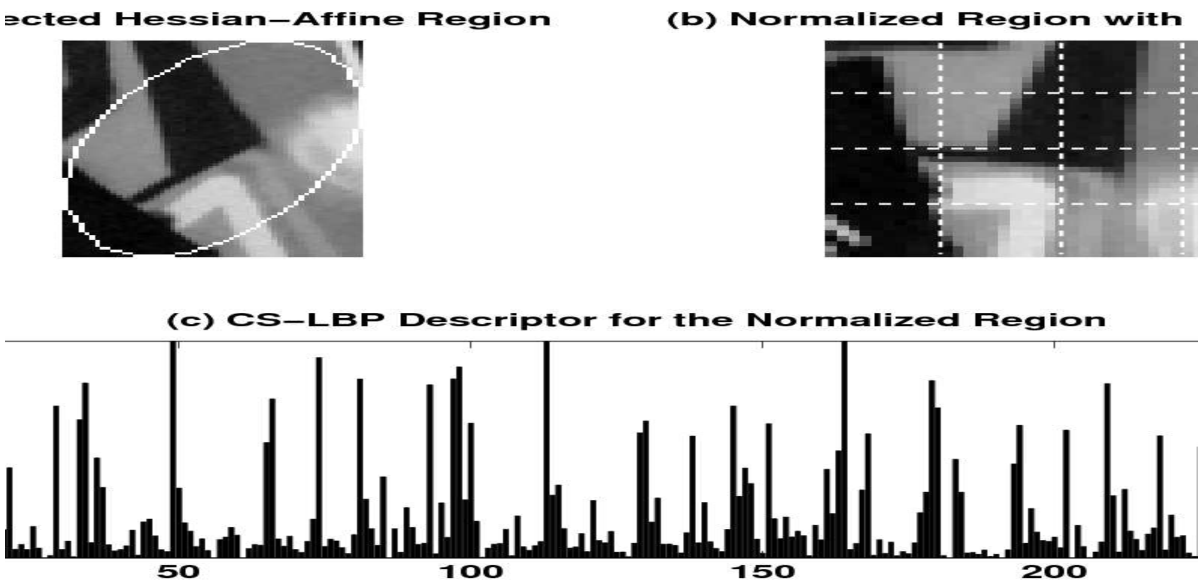
Binary Pattern

$\underline{\text{LBP}}$ =
$s(n0 - nc)2^0$ +
$s(n1 - nc)2^1$ +
$s(n2 - nc)2^2$ +
$s(n3 - nc)2^3$ +
$s(n4 - nc)2^4$ +
$s(n5 - nc)2^5$ +
$s(n6 - nc)2^6$ +
$s(n7 - nc)2^7$

$\underline{\text{CS-LBP}}$ =
$s(n0 - n4)2^0$ +
$s(n1 - n5)2^1$ +
$s(n2 - n6)2^2$ +
$s(n3 - n7)2^3$

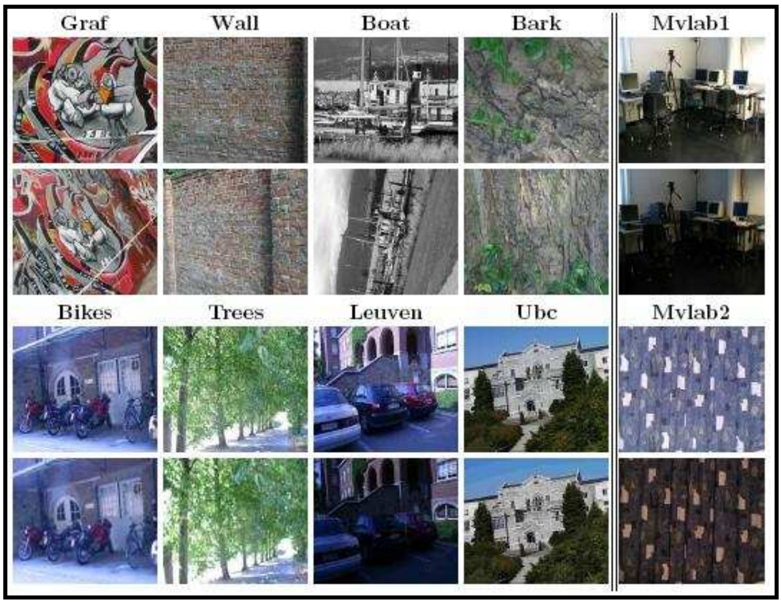
Description of interest regions





Infotech Oulu
MACHINE VISION GROUP
UNIVERSITY of OULU
OULUN YLIOPISTO

Setup for image matching experiments



• CS-LBP perfomed better than SIFT in image maching and categorization experiments, especially for images with Illumination variations

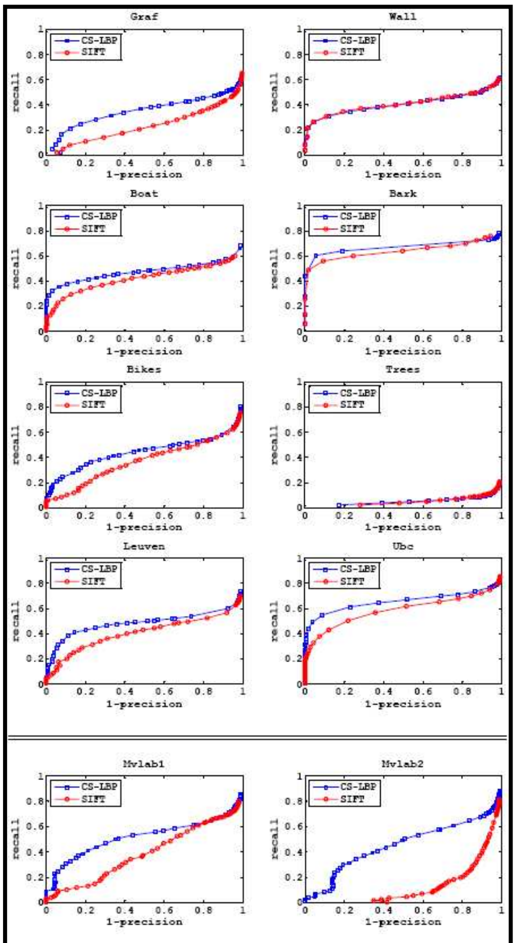


Fig. 6. Comparison results for Hessian-Affine regions.

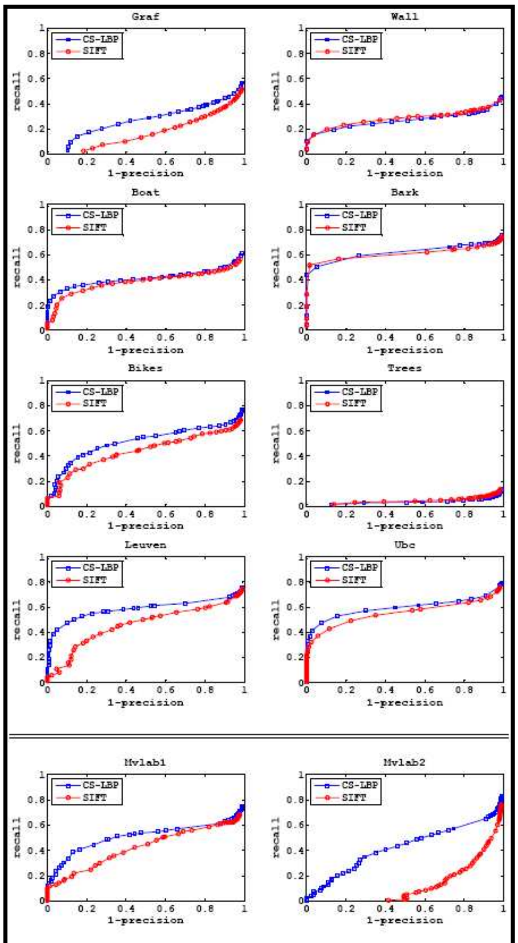
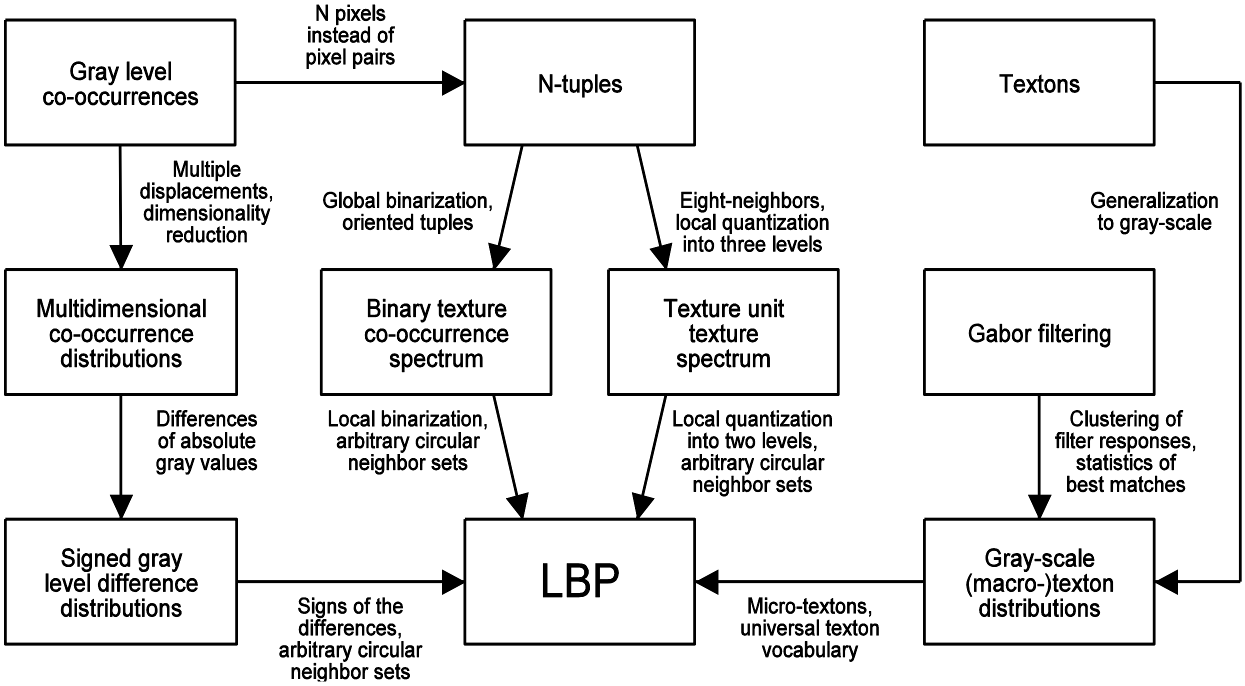


Fig. 7. Comparison results for Harris-Affine regions.

LBP unifies statistical and structural approaches



Dynamic textures (R Nelson & R Polana: IUW, 1992; M Szummer & R Picard: ICIP, 1995; G Doretto et al., IJCV, 2003)

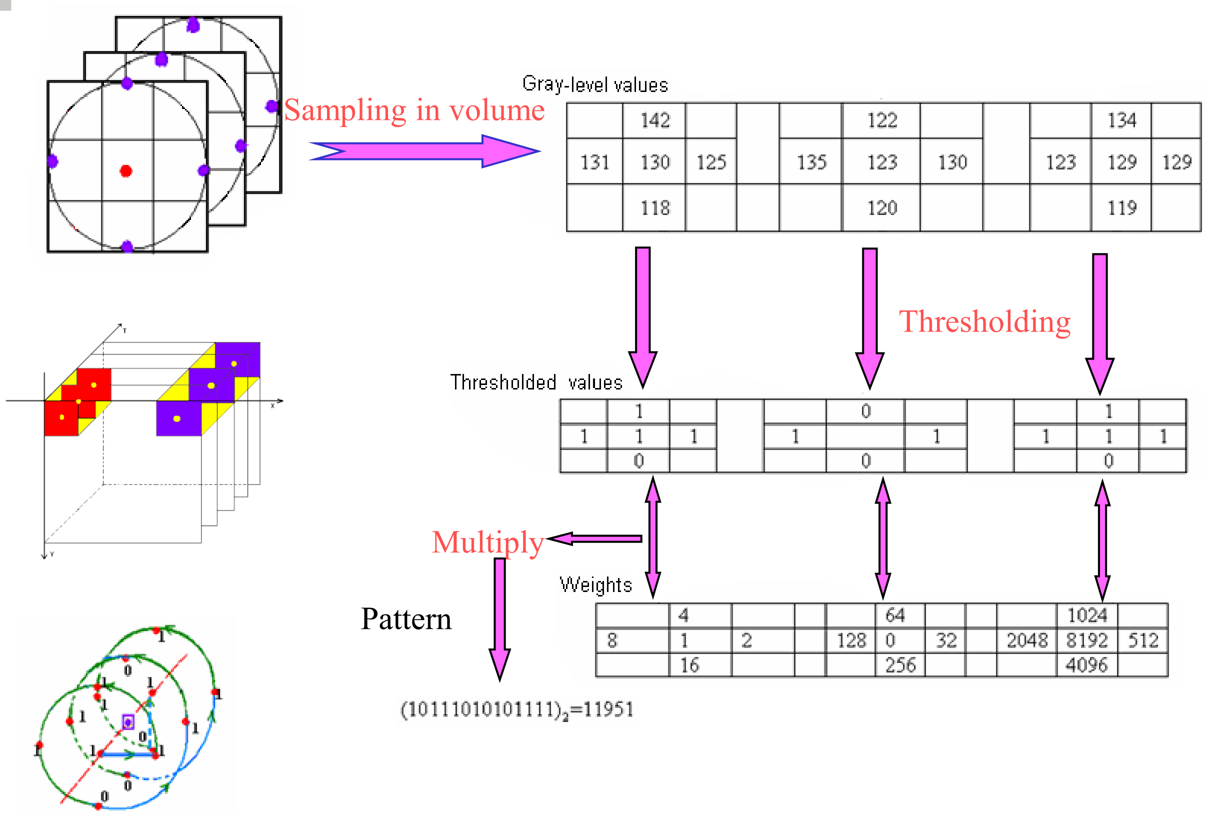
Dynamic texture recognition

Zhao G & Pietikäinen M (2007) Dynamic texture recognition using local binary patterns with an application to facial expressions. IEEE Transactions on Pattern Analysis and Machine Intelligence 29(6):915-928. (parts of this were earlier presented at ECCV 2006 Workshop on Dynamical Vision and ICPR 2006)

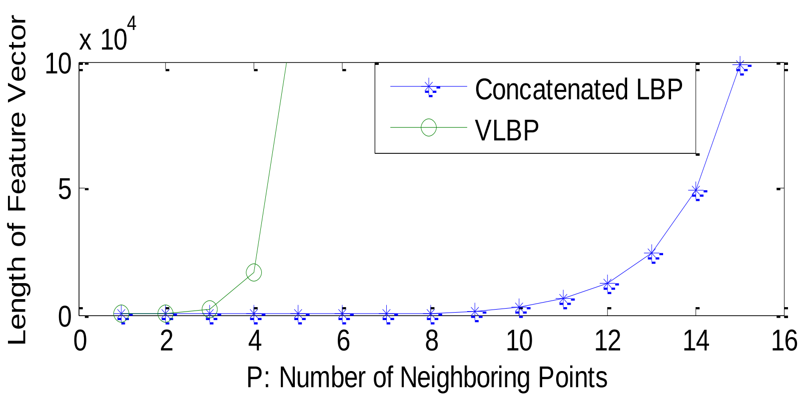
Dynamic textures

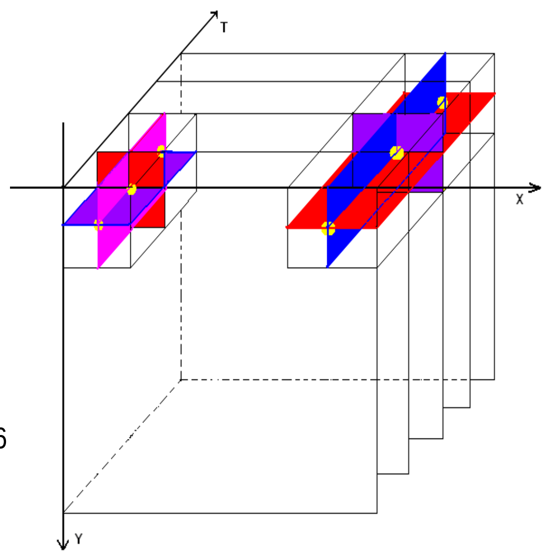
- An extension of texture to the temporal domain
- Encompass the class of video sequences that exhibit some stationary properties in time
- Dynamic textures offer a new approach to motion analysis
  - general constraints of motion analysis (i.e. scene is Lambertian, rigid and static) can be relaxed [Vidal et al., CVPR 2005]

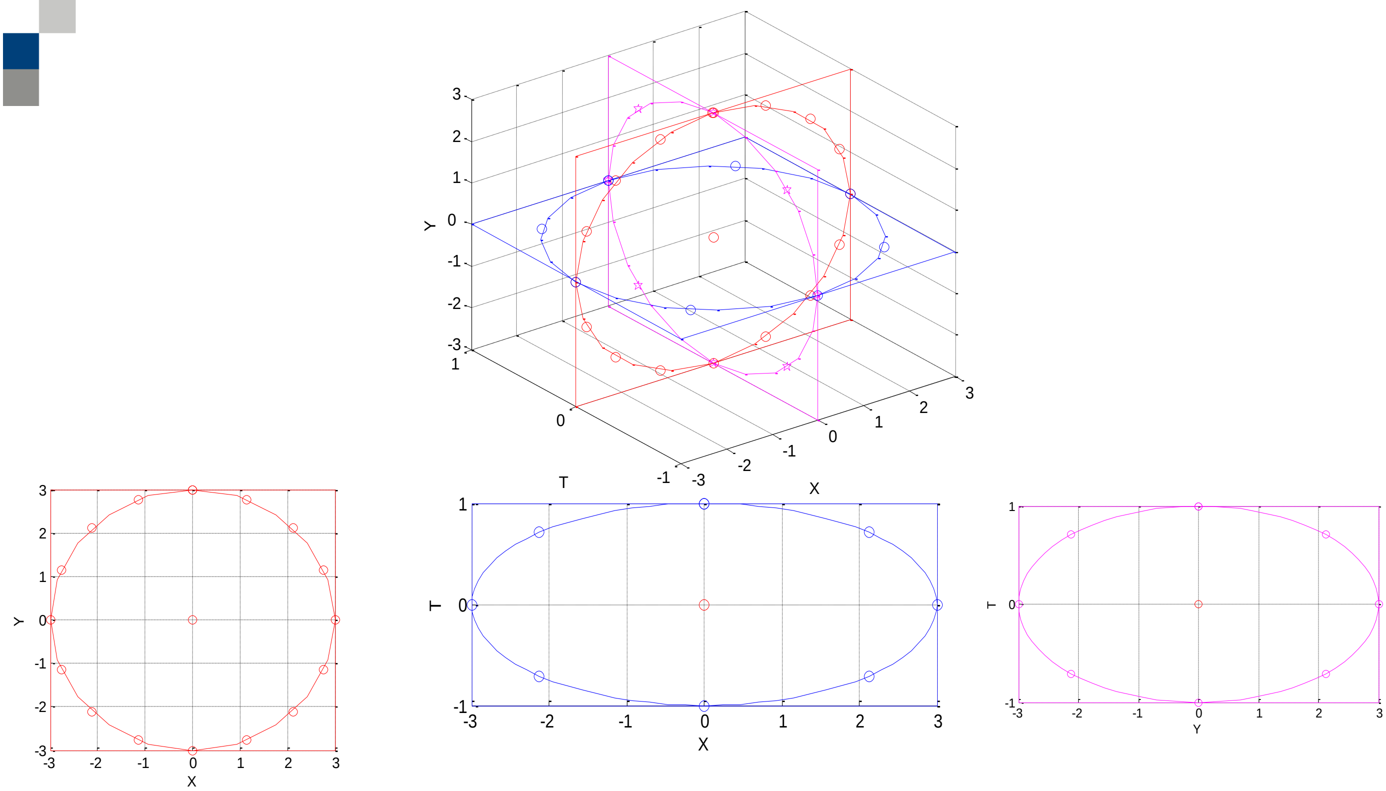
Volume Local Binary Patterns (VLBP)



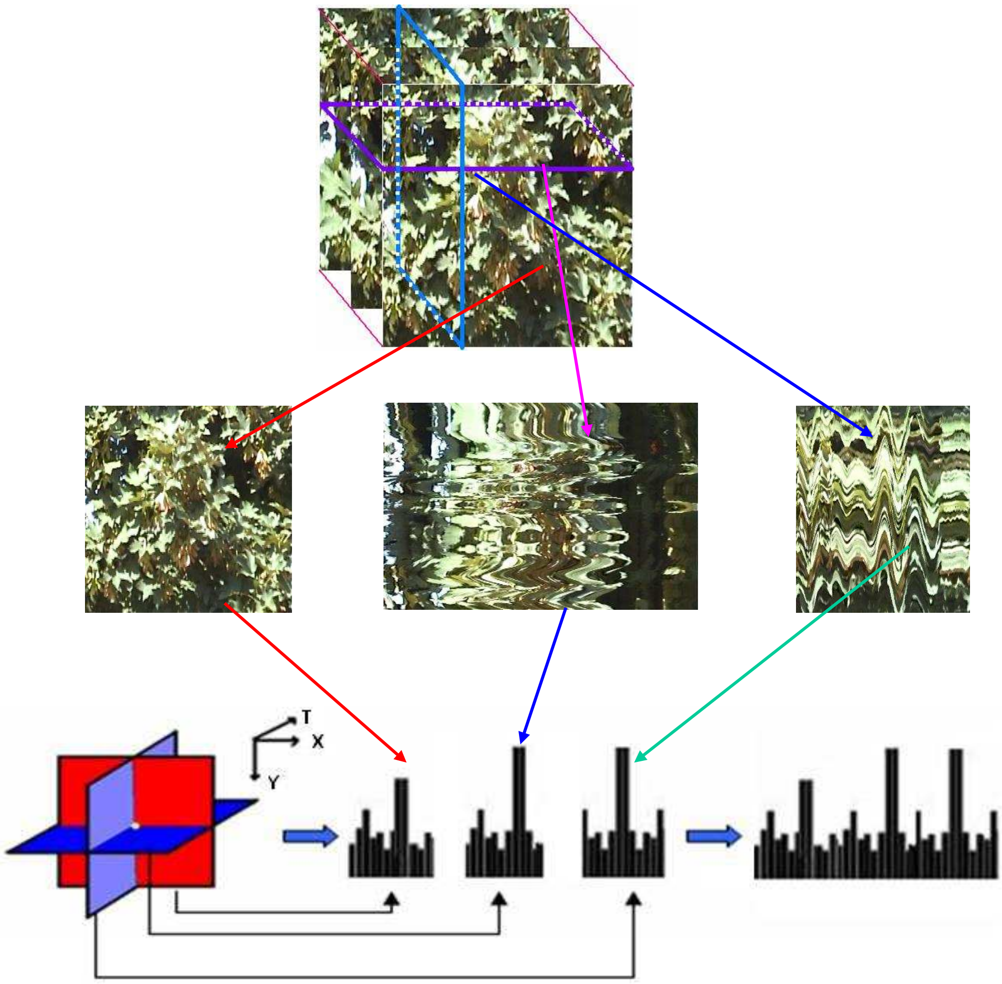
LBP from Three Orthogonal Planes (LBP-TOP)





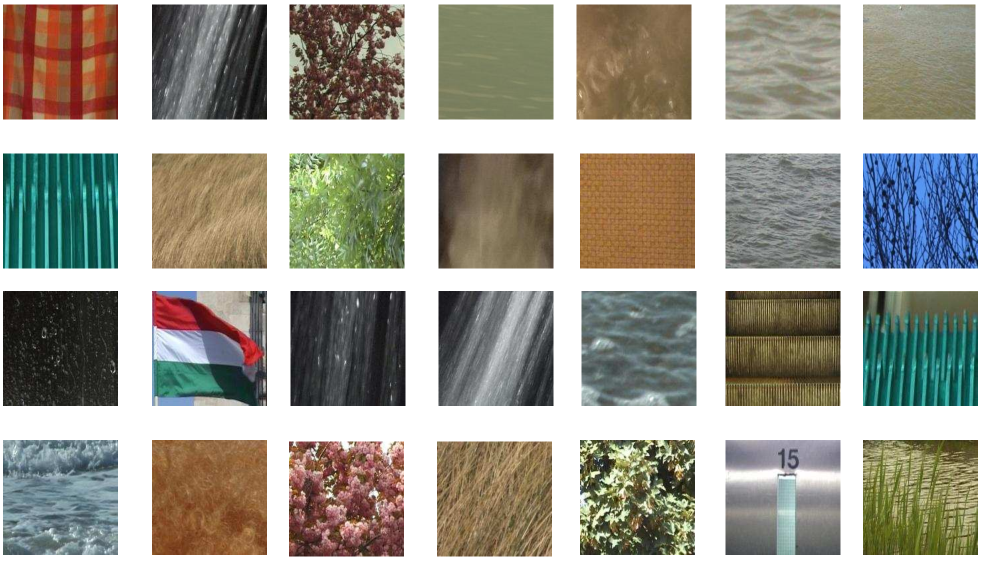


LBP-TOP



Infotech Oulu

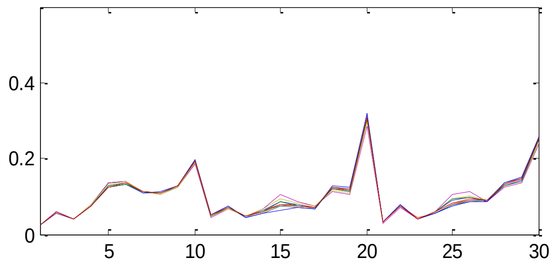
DynTex database

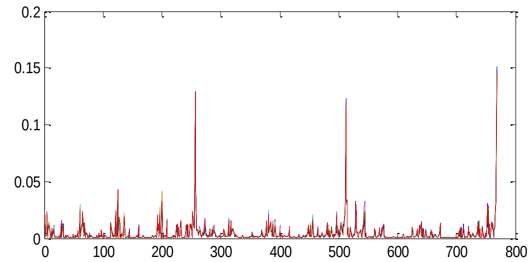


• Our methods outperformed the state-of-the-art in experiments with DynTex and MIT dynamic texture databases

Infotech Oulu
MACHINE VISION GROUP
UNIVERSITY of OULU
OULUN YLIOPISTO

Results of LBP from three planes





LBP	XY	XZ	YZ	Con	weighted
8,8,8,1,1,1 riu2	88.57	84.57	86.29	93.14	93.43[2,1,1]
8,8,8,1,1,1 u2	92.86	88.86	89.43	94.57	96.29[4,1,1]
8,8,8,1,1,1 Basic	95.14	90.86	90	95.43	97.14[5,1,2]
8,8,8,3,3,3 Basic	90	91.17	94.86	95.71	96.57[1,1,4]
8,8,8,3,3,1 Basic	89.71	91.14	92.57	94.57	95.71[2,1,8]

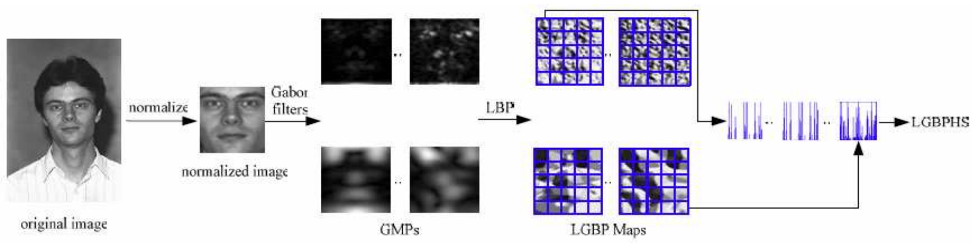
Part 2: Some recent variants of LBP

Matti Pietikäinen

Infotech Oulu
MACHINE VISION GROUP
UNIVERSITY of OULU
OULUN YLIOPISTO

Preprocessing prior to LBP feature extraction

- Gabor filtering for extracting more macroscopic information [Zhang et al., ICCV 2005]



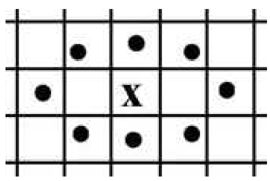
- Preprocessing for illumination normalization [Tan & Triggs, AMFG 2007]

- Edge dection
  - used to enhance the gradient information

Neighborhood topology

- Square or circular neighborhood is normally used
  - circular neighborhood important for rotation-invariant operators

- Anisisotropic neighborhood (e.g. elliptic)
  - improved results in face recognition [Liao & Chung, ACCV 2007, and in medical image analysis [Nanni et al., Artif. Intell. Med. 2010]



- Encoding similarities between patches of pixels [Wolf et al., ECCV 2008]
  - they characterize well topological structural information of face appearance

Thresholding and encoding

- Using mean or median of the neihborhood for thresholding
- Using a non-zero threshold [Heikkilä et al., IEEE PAMI 2006]
- Local tenary patterns - encoding by 3 values [Tan & Triggs, AMGF 2007]
- Extended quinary patterns – encoding by 4 values [Nanni et al., Artif. Intell. Med. 2010]
- Soft LBP [Ahonen & Pietikäinen, Finsig 2007]
- Scale invariant local ternary pattern [Liao et al., CVPR 2010]
  - for background subtraction applications

Robust LBP [Heikkilä et al., PAMI 2006]

The value of the LBP code of a pixel (x_c, y_c) is given by:

$$LBP_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c)2^p \qquad s(x) = \begin{cases} 1, if\ x \geq 0; \\ 0, otherwise. \end{cases}$$

In robust LBP, the term $s(g_p – g_c)$ is replaced with $s(g_p – g_c + a)$

Allows bigger changes in pixel values without affecting thresholding results
- improved results in background subtraction [Heikkilä et al., PAMI 2006]
- was also used in CS-LBP interest region descriptor [Heikkilä et al., PR 2009]

Local ternary patterns (LTP) [Tan & Triggs, AMGF 2007]

Ternary code:1100(-1)(-1)00

1	1	0
0		0
0	-1	-1

Binary code: 11000000

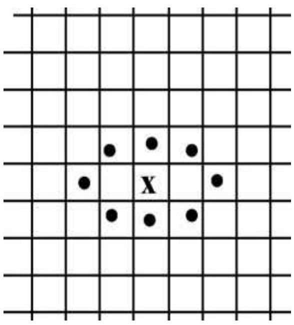
1	1	0
0		0
0	0	0

Binary code: 00001100

0	0	0
0		0
0	1	1

Enlongated quinary patterns (EPQ) [Nanni et al., Artif. Intell. Med. 2010]

- Different ellipse orientations can be used to optimize the performance



1	-2	1
-1		2
1	-1	0

0	0	0
0		1
0	0	0

1	0	1
0		0
1	0	0

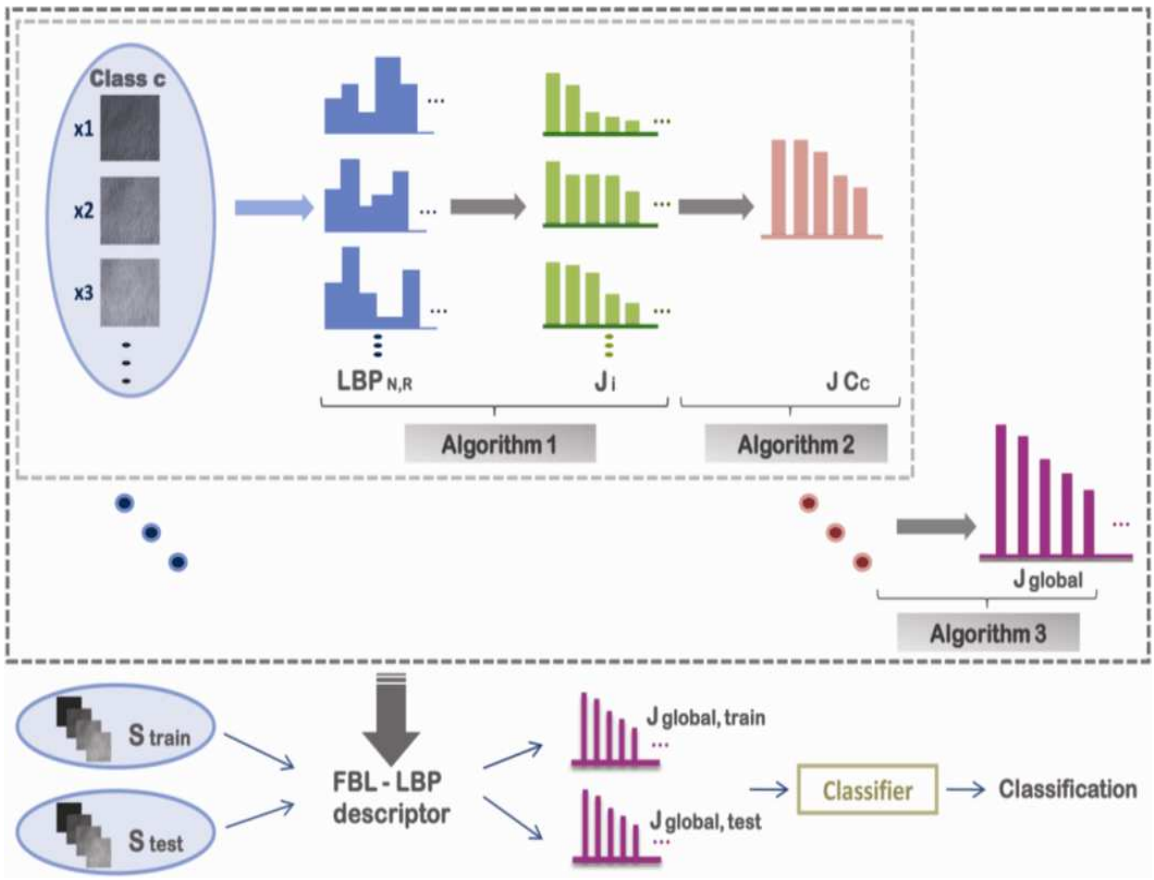
0	0	0
1		0
0	1	0

0	1	0
0		0
0	0	0

Feature selection and learning

- Feature selection e.g. with AdaBoost to reduce the number of bins [Zhang et al., LNCS 2005]
- Subspace methods projecting LBP features into a lower-dimensional space [Shan et al., ICPR 2006], [Chan et al., ICB 2007]
- Learning the most dominant and discriminative patterns [Liao et al., IEEE TIP 2009], [Guo et al., ACCV 2010]

A learning-based LBP using Fisher separation criterion [Guo et al., ACCV 2010]



Use of complementary contrast/magnitude information

- LBP was designed as a complementary measure of local contrast, using joint LBP/C or LBP/VAR histograms
- LBPV puts the local contrast into 1-dimensional histogram [Guo et al., Pattern Recogn. 2010]
- Completed LBP (CLBP) considers complementary sign and magnitude vectors [Guo et al., IEEE TIP 2010]
- Weber law descriptor (WLD) includes excitation and orientation components [Chen et al., IEEE PAMI 2010]

Completed LBP [Guo et al., IEEE TIP 2010]

a) 3 x 3 sample block
b) The local differences
c) The sign component
d) The magnitude component

9	12	34
10	**25**	28
99	64	56

(a)

-16	-13	9
-15		3
74	39	31

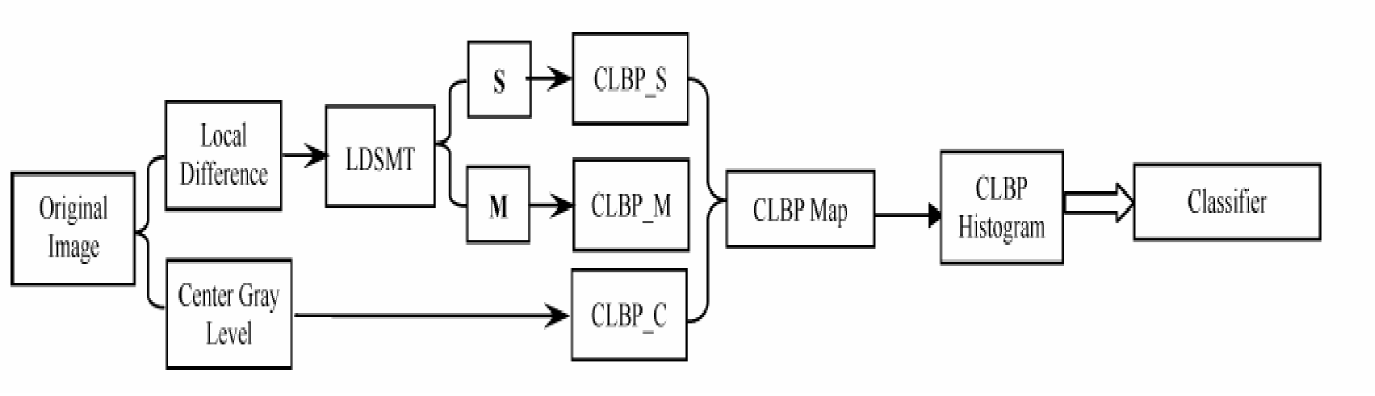
(b)

-1	-1	1
-1		1
1	1	1

(c)

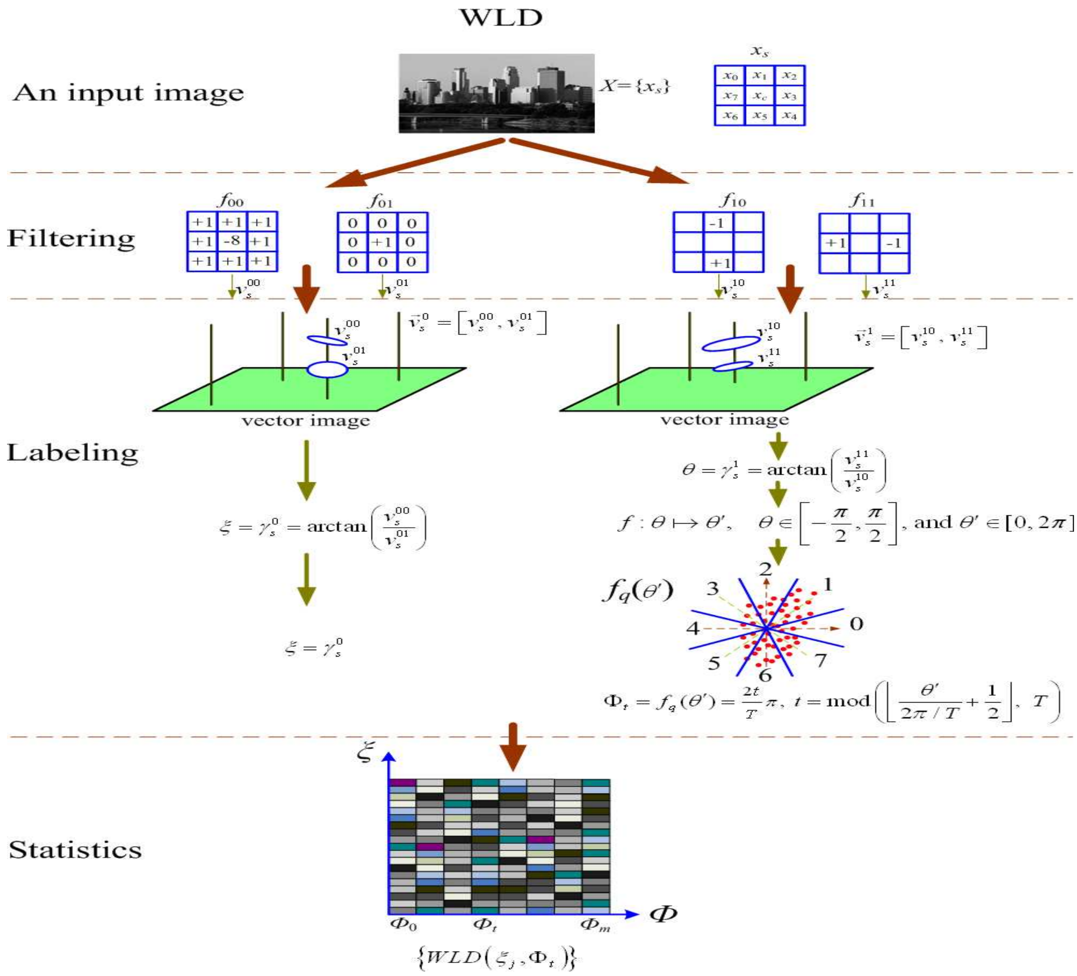
16	13	9
15		3
74	39	31

(d)



WLD: Weber law descriptor [Chen et al., PAMI 2010]

Composed of excitation and orientation components



Infotech Oulu

Local Phase Quantization (LPQ)

Janne Heikkilä

Tutorial: Image and Video Description with Local Binary Pattern Variants

Motivation

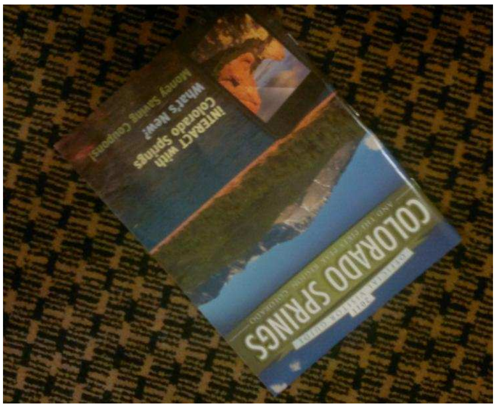
- Local descriptors are widely used in image analysis
  - Interest point descriptors (e.g. SIFT, SURF)
  - Texture descriptors (e.g. LBP, Gabor texture features)
- Descriptors should be robust to various degradations including geometric distortions, illumination changes and blur.
- Blur-sensitivity of the descriptors has not been much studied in the literature.
- Sharp images are often assumed.
- We propose using Fourier phase information for blur insensitive texture analysis.

Previous work on phase based methods

- Phase correlation [Kuglin & Hines, ICCS 1975] has been successfully used for image registration.
- Eklundh [CGIP 1979] suggested that ”Fourier features based on the phase rather than the amplitude do not seem to be useful for texture classification”.
- Oppenheimer & Lim [Proc. IEEE 1981] showed the importance of the phase information in signal reconstruction.
- Daugman [PAMI 1993] used phase-quandrant coding of multi-scale 2-D Gabor wavelet coefficients for iris recognition.
- Fischer & Bigün [SCIA 1995] presented a method for texture bondary tracking based on Gabor phase.
- Zhou et al. [ICIP 2001] introduced a texture feature based on a histogram of local Fourier coefficients (magnitude and phase).
- Zhang et al. [TIP 2007] proposed histogram of Gabor phase patterns (HGPP) for face recognition.

Image blur

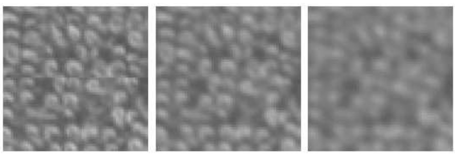
- Blur is typically caused by defocus or motion.
- Blur is usually harmful for image analysis.



- Two approaches:
  – restoration (deblurring)
  – using blur-insensitive features

Performance of some texture analysis methods under blur

- Outex database with artificially generated circular blur.
- Examples: blur with radius r=0,1,2.

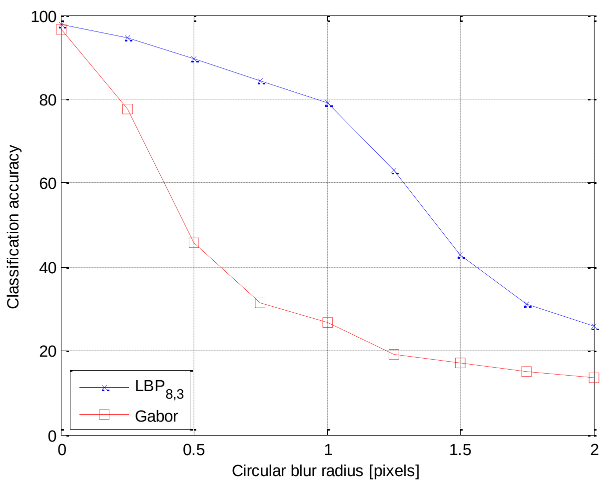


- Methods:

  LBP: [Ojala et al., PAMI 2002]

  Gabor: [Manjunath & Ma, PAMI 1996]

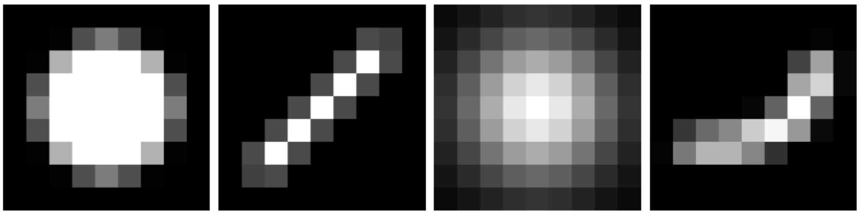
- Results



- Significant drop-off in the accuracy!

Blur model

- Image blur can be modelled by
$$g(\mathbf{x}) = (f * h)(\mathbf{x})$$
where f is the image and h is the PSF.
- Examples of discrete PSFs: (a) out of focus blur, (b) linear motion blur, (c) Gaussian blur, and (d) arbitrary blur.



(a) (b) (c) (d)

- In many cases the PSF is centrally symmetric.

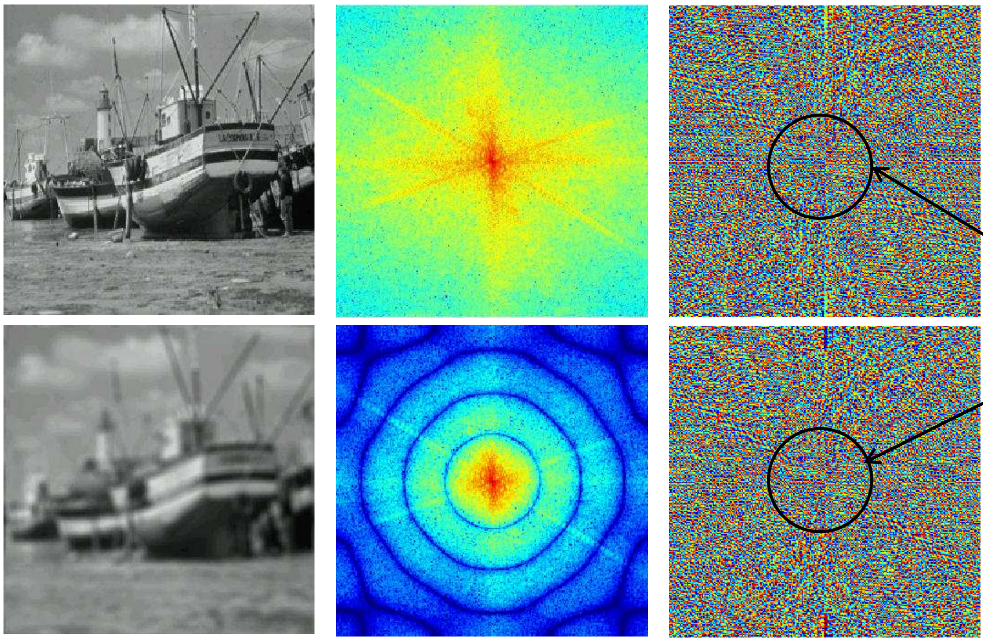
Blur invariance of phase spectrum

- In frequency domain the blur model corresponds to
$G(\mathbf{u}) = F(\mathbf{u}) \cdot H(\mathbf{u})$ or
$$\begin{aligned} |G(\mathbf{u})| &= |F(\mathbf{u})| \cdot |H(\mathbf{u})| \quad \text{and} \\ \angle G(\mathbf{u}) &= \angle F(\mathbf{u}) + \angle H(\mathbf{u})\,. \end{aligned}$$
- For centrally symmetric blur
$$\angle H(\mathbf{u}) = \begin{cases} 0 & \text{if } H(\mathbf{u}) \geq 0 \\ \pi & \text{if } H(\mathbf{u}) < 0 \end{cases}$$
- We notice that
$$\angle G(\mathbf{u}) = \angle F(\mathbf{u}) \quad \text{for all } H(\mathbf{u}) \geq 0$$

Blur invariance of phase spectrum

Original

Blurred (circular blur d=4)

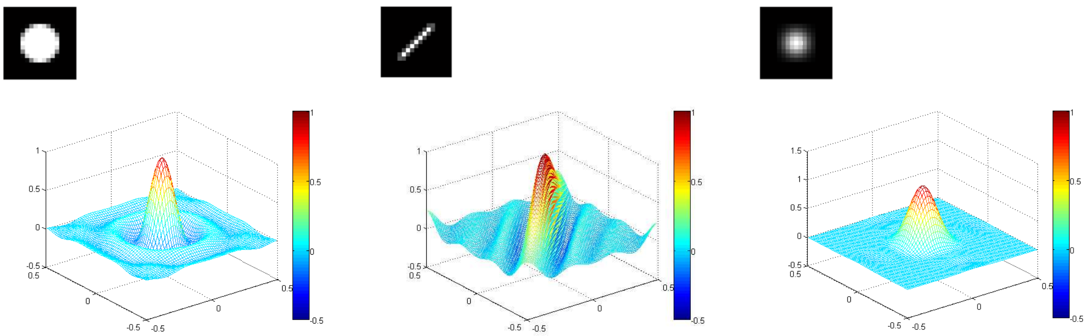


Phase remains unchanged

image amplitude spectrum phase spectrum

What frequencies should be used?

- Frequency responses of 3 centrally symmetric PSFs:



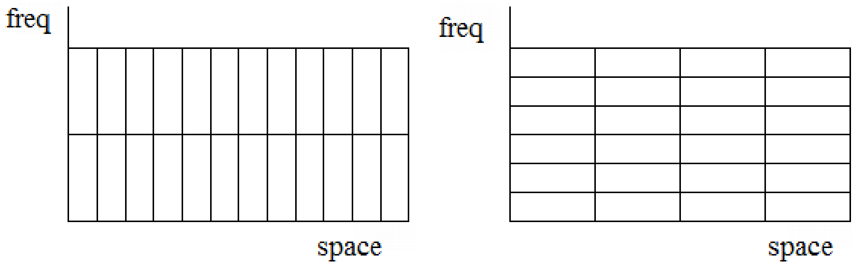
- For low frequencies typically $H(\mathbf{u}) \geq 0$
  $\rightarrow$ It is safe to select low frequency phase angles as blur invariants.
- Low frequency components carry most of the information.

MACHINE VISION GROUP

UNIVERSITY of OULU
OULUN YLIOPISTO

From global to local

- For local descriptors it is necessary to compute the phase locally.
- Local frequency characteristics can be extracted using frequency selective filters.
- Higher spatial resolution implies lower frequency resolution and vice versa.



- Local phase cannot be measured accurately!
- Blur does not have to be spatially invariant.

Computing the local phase

- A set of frequency selective filters is needed. Selection criteria:
  - Low spatial frequency (guarantees that $H(\mathbf{u}) \geq 0$)
  - Narrow band (less distortion)
- The phase is derived from the analytic signal

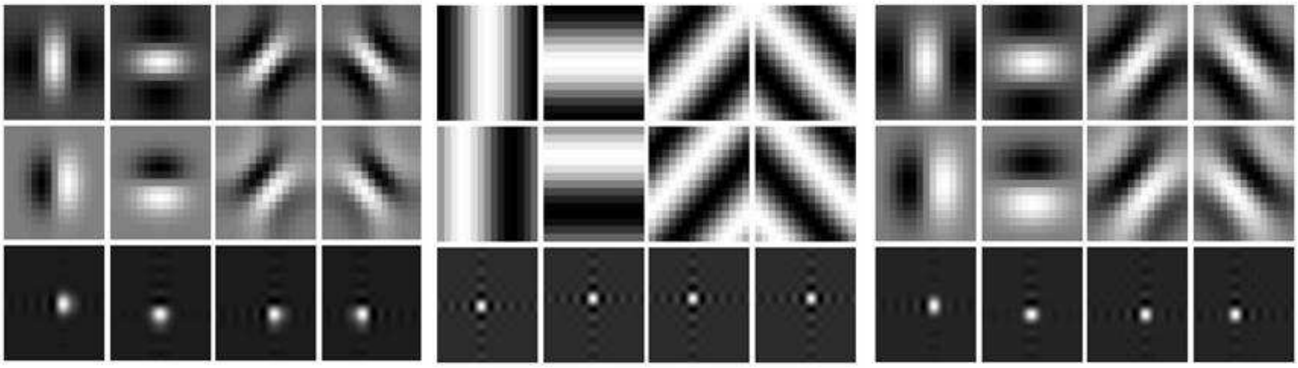
$$f_A(x) = f(x) - i\, f_H(x)$$

  where $f_H(x)$ is the Hilbert transform of $f(x)$
- Filters are complex valued (real and imaginary parts).
- 1-D case: quadrature filters (Gabor, log-Gabor, derivative of Gaussian, Cauchy, etc.)
- 2-D case: not well-defined, many alternatives:
  - Monogenic signal (isotropic)
  - Directional 2-D quadrature filters
  - Short-term Fourier transform (STFT)

Examples of complex valued filters

Four 15 by15 filters with lowest non-zero frequency components:

Gaussian derivative   STFT (uniform window)   STFT (Gaussian window)



Row 1: real part,
Row 2: imaginary part,
Row 3: frequency response

Notice: Filters are strongly truncated!

Short-Term Fourier Transform (STFT)

- Discrete version of the STFT is defined by

$$F^S(\mathbf{k},\mathbf{u}) = \sum_{\mathbf{l}\in\mathbb{Z}^2} f(\mathbf{l})w(\mathbf{k}-\mathbf{l})e^{-j2\pi(\mathbf{k}-\mathbf{l})^T\mathbf{u}}$$

where **k** is the position and **u** is the 2-D frequency.

- Various alternatives exist for the window function *w*(**x**), for example (*m* is the window size):

$$w^U(\mathbf{x}) = \begin{cases} 1 & ||\mathbf{x}||_\infty \leq m/2 \\ 0 & \text{otherwise,} \end{cases}$$

$$w^G(\mathbf{x}) = \begin{cases} e^{-\frac{1}{2\sigma_S^2}\mathbf{x}^T\mathbf{x}} & ||\mathbf{x}||_\infty \leq m/2 \\ 0 & \text{otherwise,} \end{cases}$$ (Gabor filter)

- STFT is separable
  - It can be implemented with 1-D convolutions.

Quantization of the phase angle

- Phase angles could be used directly but it would result in long feature vectors.
- Quantizing the phase to four angles [0, $\pi/2$, π, $3\pi/2$] causes only moderate distortion.

original

phase quantized

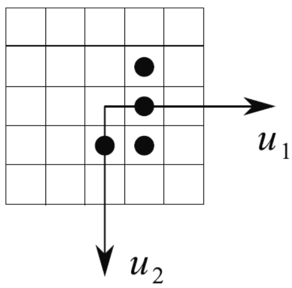
Local Phase Quantization (LPQ)

- The local frequency coefficients $F(\mathbf{k}, \mathbf{u})$ are computed for all pixels at some frequencies $\mathbf{u} \in \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_L\}$.
- We use the following frequencies (L = 4):

$$\mathbf{u}_1 = [a, 0]^T,\ \mathbf{u}_2 = [0, a]^T,\ \mathbf{u}_3 = [a, a]^T, \text{ and}$$

$$\mathbf{u}_4 = [a, -a]^T$$

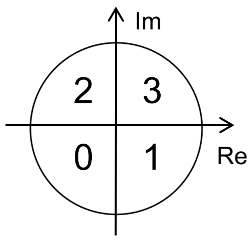
where $a = 1/m$.



- The coefficients are quantized using:

$$Q(F(\mathbf{k}, \mathbf{u})) = \operatorname{sgn}(\operatorname{Re}\{F(\mathbf{k}, \mathbf{u})\}) + 2\operatorname{sgn}(\operatorname{Im}\{F(\mathbf{k}, \mathbf{u})\})$$

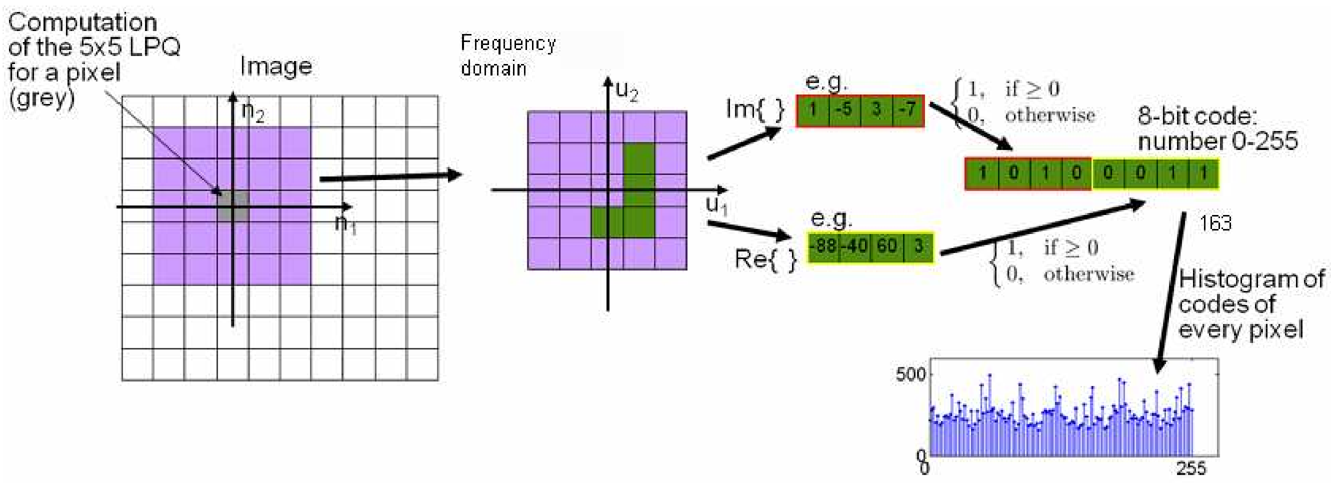
that results in a two-bit code.



- An 8-bit codeword is obtained from 4 coefficients.
- Codewords are histogrammed and used as a feature vector (LPQ descriptor).

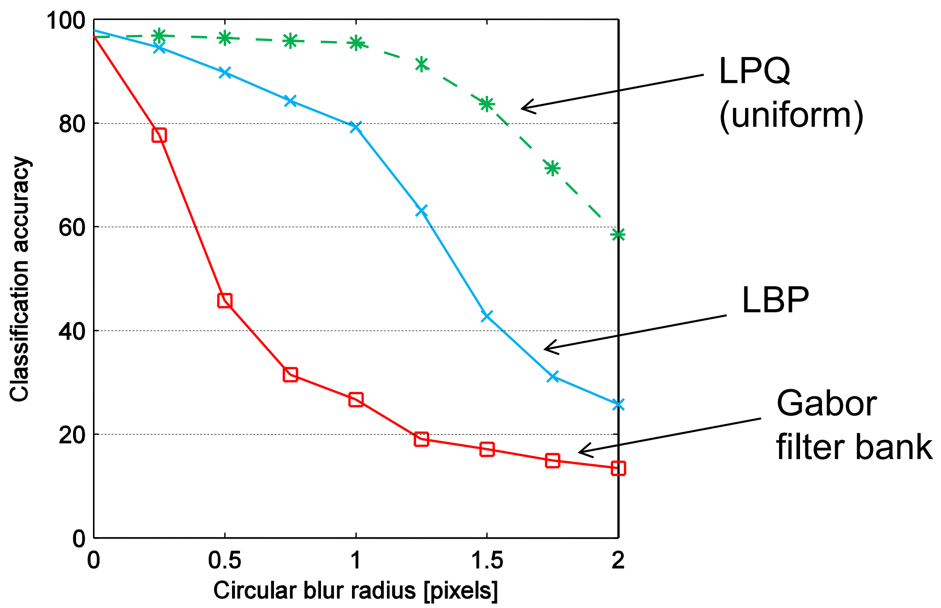
Basic LPQ descriptor

- Illustration of the algorithm:



Experimental results (1)

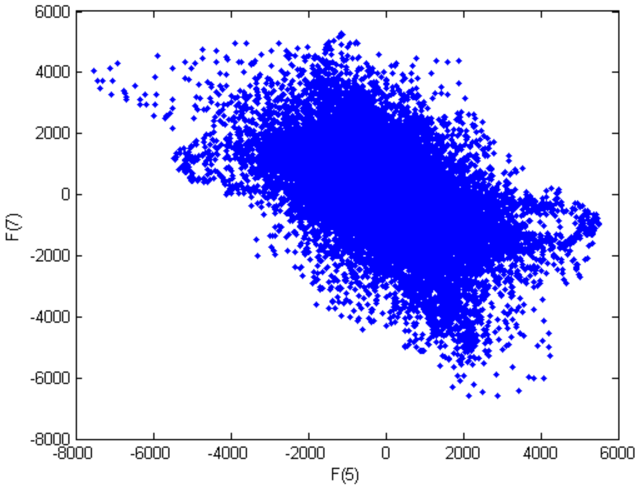
- Classification with artificially blurred textures (Outex TC 00001):



m = 7

Can we do better?

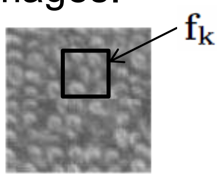
- The filter responses are correlating. Example: responses of filters 5 and 7:



- Scalar quantization is efficient only if the samples are independent.
- Performance can be improved using decorrelation (PCA).

Image model

- Neighboring pixels are strongly correlated in natural images.
- Let $\mathbf{l}_1, \mathbf{l}_2, \ldots, \mathbf{l}_{m^2}$ denote the pixel positions in an *m* by *m* image patch



$$\mathbf{f_k} = [f(\mathbf{l}_1 + \mathbf{k}), f_\mathbf{k}(\mathbf{l}_2 + \mathbf{k}), \cdots, f_\mathbf{k}(\mathbf{l}_{m^2} + \mathbf{k})]^T$$

- Pixels in $\mathbf{f_k}$ are considered as realizations of a random process.
- Correlation coefficient between adjacent pixels is denoted by ρ.
- Covariance between two positions $\mathbf{l}_i$ and $\mathbf{l}_j$ is assumed to be exponentially related to their Euclidean distance so that $\sigma_{ij} = \rho^{||\mathbf{l}_i - \mathbf{l}_j||}$. **Notice:** blur is now assumed to be isotropic!
- The covariance matrix of $\mathbf{f_k}$ can be expressed as

$$\mathbf{C} = \begin{bmatrix} 1 & \sigma_{1,2} & \cdots & \sigma_{1,m^2} \\ \sigma_{2,1} & 1 & \cdots & \sigma_{2,m^2} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{m^2,1} & \sigma_{m^2,2} & \cdots & 1 \end{bmatrix}$$

- Covariance matric could be also learned from training data.

STFT revisited

- Using vector notation we can rewrite the STFT as
$$F(\mathbf{k}, \mathbf{u}) = \psi_{\mathbf{u}}^T \mathbf{f}_{\mathbf{k}}$$
where
$$\psi_{\mathbf{u}}^T = [w(\mathbf{l}_1)e^{-j2\pi \mathbf{l}_1^T \mathbf{u}}, w(\mathbf{l}_2)e^{-j2\pi \mathbf{l}_2^T \mathbf{u}}, \ldots, w(\mathbf{l}_{m^2})e^{-j2\pi \mathbf{l}_{m^2}^T \mathbf{u}}]$$
- Let us consider all the frequency samples $\mathbf{u} \in \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_L\}$.
- We separate the real and imaginary of the filters:
$$\mathbf{\Psi} = [\mathbf{\Psi}_R, \mathbf{\Psi}_I]^T$$
where
$$\mathbf{\Psi}_R = \mathrm{Re}\{[\psi_{\mathbf{u}_1}, \psi_{\mathbf{u}_2}, \cdots, \psi_{\mathbf{u}_L}]\}$$
and
$$\mathbf{\Psi}_I = \mathrm{Im}\{[\psi_{\mathbf{u}_1}, \psi_{\mathbf{u}_2}, \cdots, \psi_{\mathbf{u}_L}]\}$$

STFT revisited (2)

- The frequency samples and the pixel values have linear dependence:
$$\mathbf{F}_{\mathbf{k}} = \mathbf{\Psi} \mathbf{f}_{\mathbf{k}}$$
where
$$\mathbf{F}_{\mathbf{k}} = [\mathbf{F}_{\mathbf{k},R}, \mathbf{F}_{\mathbf{k},I}]^T$$
and
$$\mathbf{F}_{\mathbf{k},I} = \mathrm{Im}\{[F(\mathbf{k}, \mathbf{u}_1), F(\mathbf{k}, \mathbf{u}_2), \cdots, F(\mathbf{k}, \mathbf{u}_L)]\}$$
$$\mathbf{F}_{\mathbf{k},R} = \mathrm{Re}\{[F(\mathbf{k}, \mathbf{u}_1), F(\mathbf{k}, \mathbf{u}_2), \cdots, F(\mathbf{k}, \mathbf{u}_L)]\}$$

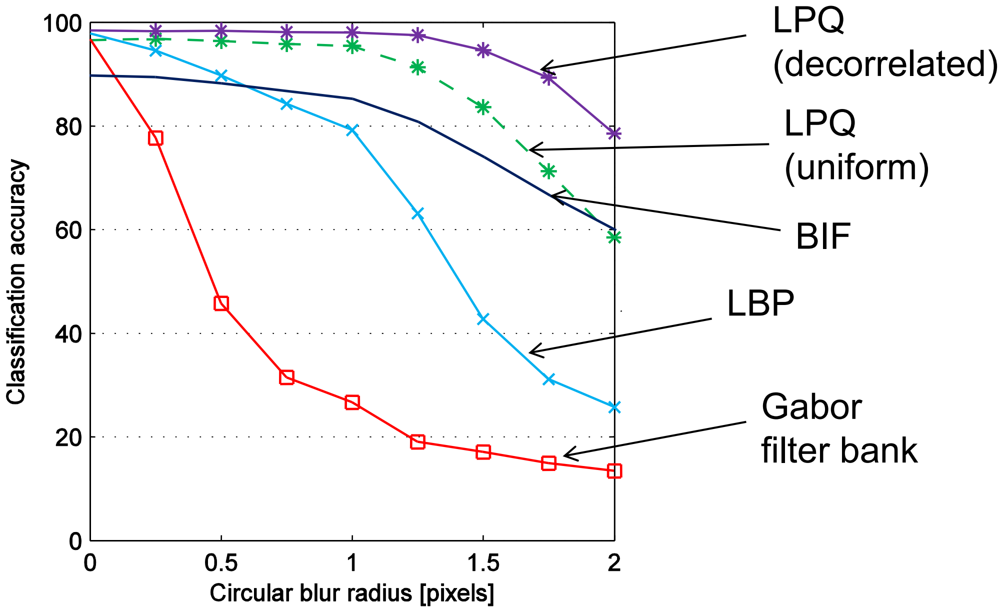
Decorrelated LPQ

- The covariance matrix of the frequency samples becomes
$$\mathbf{D} = \mathbf{\Psi} \mathbf{C} \mathbf{\Psi}^T$$
- If L = 4, **D** is an 8 by 8 matrix.
- We can employ whitening transform
$$\mathbf{G_k} = \mathbf{V}^T \mathbf{F_k}$$
where **V** is an orthonormal matrix derived by using SVD:
$$\mathbf{D} = \mathbf{U \Sigma V}^T$$
- The vector $\mathbf{G_k}$ is quantized
$$q_j = \begin{cases} 1, & \text{if } g_j \geq 0 \\ 0, & \text{otherwise} \end{cases}$$
- 8-bit integers obtained are used in the same way as in the basic LPQ.

Reference: Ojansivu V & Heikkilä J (2008) Blur insensitive texture classification using local phase quantization. Image and Signal Processing, ICISP 2008 Proceedings, Lecture Notes in Computer Science 5099:236-243.

Experimental results (2)

- Classification with artificially blurred textures (Outex TC 00001):



m = 7,
ρ = 0.9

Comparison with other texture descriptors

	LPQ	BIF [Grosier & Griffin, CVPR 2008]	VZ-joint [Varma & Zisserman, CVPR 2003]	VZMR8 [Varma & Zisserman, IJCV 2005]
Filter bank	yes*	yes	no	yes
Number of filters	8	24	-	38
Multiscale	no	yes	no	yes
Rotation invariant	no**	yes	no**	yes
Codebook	fixed	fixed	learned	learned
Histogram length	256	1296	610***	610***

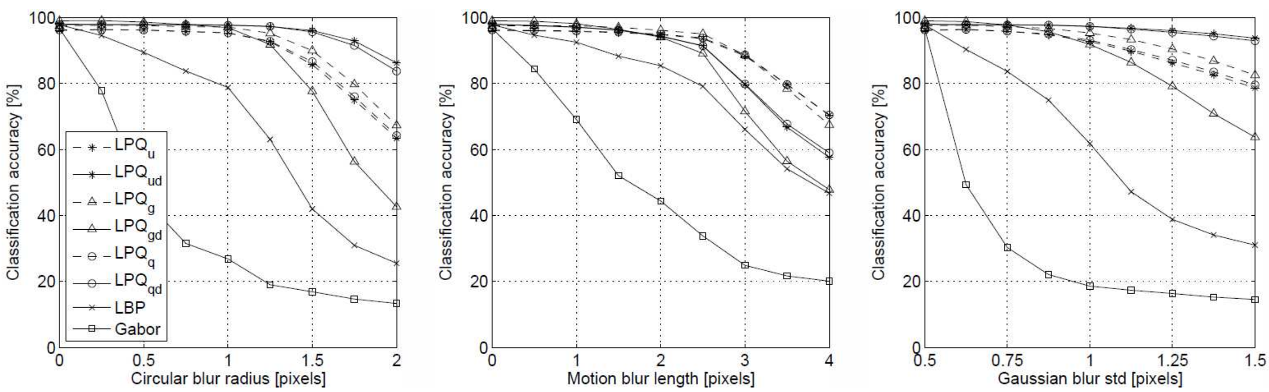
* truncated filters

** inherently not rotation invariant

*** with the CUReT dataset

Results with different filters

- Three types of filters (m = 7) were tested
  - STFT using uniform weighting with and without decorrelation
  - STFT using Gaussian weighting (Gabor filter) w & w/o decorrelation.
  - Quadrature filter (derivative of Gaussian) w & w/o decorrelation.
- Outex



Results with different filters (2)

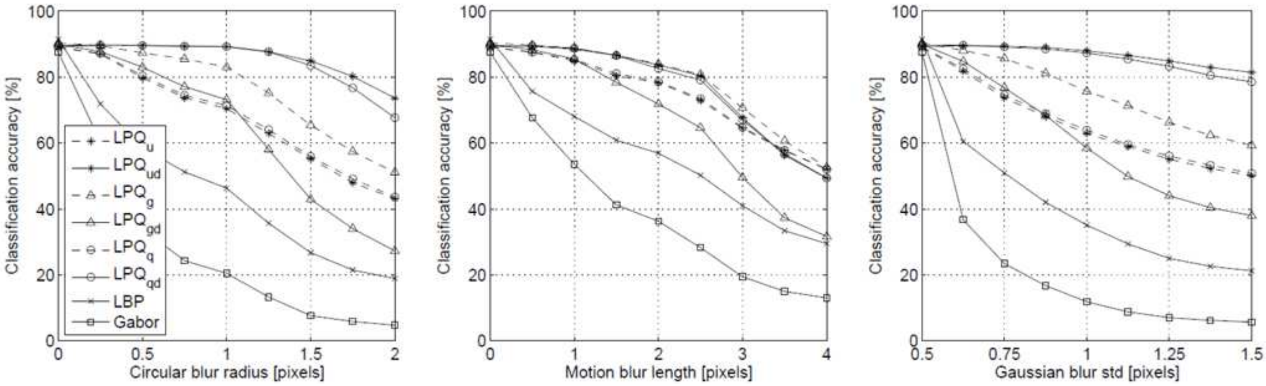
- Sharp images (Outex02, m = 3):

	LPQ_u	LPQ_{ud}	LPQ_g	LPQ_{gd}	LPQ_q	LPQ_{qd}	LBP	Gabor
Accuracy	87.4 %	92.3 %	84.9 %	**94.0 %**	88.0 %	93.4 %	90.2 %	90.2 %

- Varying filter sizes:

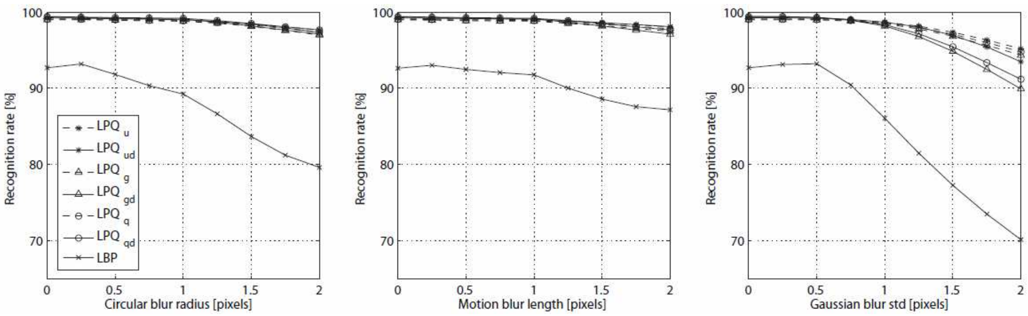
		Blur radius												
		0	0.25	0.50	0.75	1.00	1.25	1.50	1.75	2.00	2.25	2.50	2.75	3.00
m	3	98.9 %	94.0 %	78.2 %	73.5 %	70.3 %	61.9 %	45.6 %	28.1 %	20.9 %	17.3 %	13.6 %	12.8 %	12.1 %
	5	**99.1 %**	**99.2 %**	**99.2 %**	**98.9 %**	**98.6 %**	95.8 %	83.8 %	64.7 %	49.4 %	35.3 %	22.4 %	18.3 %	17.3 %
	7	97.9 %	97.9 %	97.7 %	97.8 %	97.7 %	**97.2 %**	**96.0 %**	**92.9 %**	**86.2 %**	69.5 %	51.4 %	37.7 %	31.6 %
	9	93.1 %	93.3 %	93.3 %	93.2 %	93.0 %	92.3 %	90.9 %	88.8 %	86.0 %	**78.8 %**	**67.6 %**	57.2 %	48.7 %
	11	84.7 %	84.5 %	84.4 %	84.2 %	84.1 %	83.7 %	82.6 %	81.0 %	78.3 %	73.6 %	67.3 %	**61.9 %**	**57.3 %**

- Brodatz data (m = 7):



Application: face recognition

- A modified version of [Ahonen et al., PAMI 2006].
- CMU PIE database



- Face Recognition Grand Challenge (1.0.4)

	LPQ_u	LPQ_{ud}	LPQ_g	LPQ_{gd}	LPQ_q	LPQ_{qd}	LBP
Without preprocessing	48.8 %	45.9 %	47.5 %	38.0 %	48.2 %	41.8 %	32.6 %
With preprocessing	71.9 %	74.5 %	70.4 %	68.6 %	69.2 %	65.6 %	63.8 %

Reference: Ahonen T, Rahtu E, Ojansivu V & Heikkilä J (2008) Recognition of blurred faces using local phase quantization. Proc. 19th International Conference on Pattern Recognition (ICPR 2008), Tampa, FL, 4p.

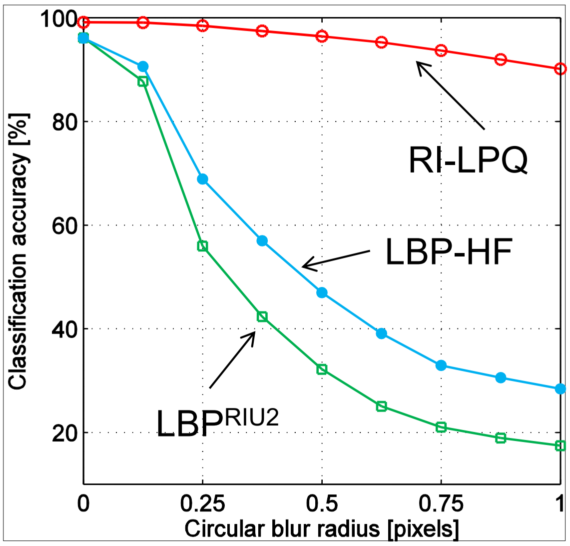
Rotation invariant LPQ

- We define blur insensitive local characteristic orientation by

$$\xi(\mathbf{k}) = \angle \sum_{i=0}^{M-1} \mathrm{sgn}(\mathrm{Im}\{F(\mathbf{k}, \mathbf{u}(\varphi_i))\}) e^{j\varphi_i}$$

where $\varphi_i = 2\pi i/M$ and M is the number of samples (we use $M = 36$).

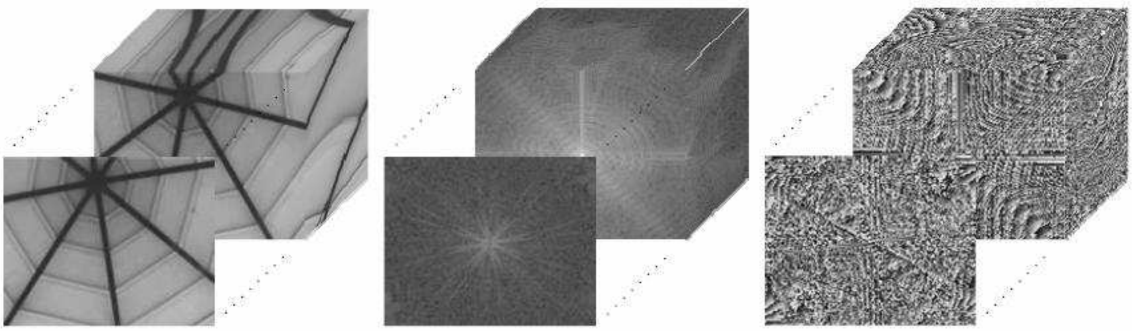
- Characteristic orientation is used to normalize the orientation locally for each location **k**.
- LPQ is computed for the normalized patches.
- Results with Outex10 (rotation and blur):



Reference: Ojansivu V, Rahtu E & Heikkilä J (2008) Rotation invariant blur insensitive texture analysis using local phase quantization. Proc. 19th International Conference on Pattern Recognition (ICPR), Tampa, FL, 4 p.

Spatio-temporal texture analysis

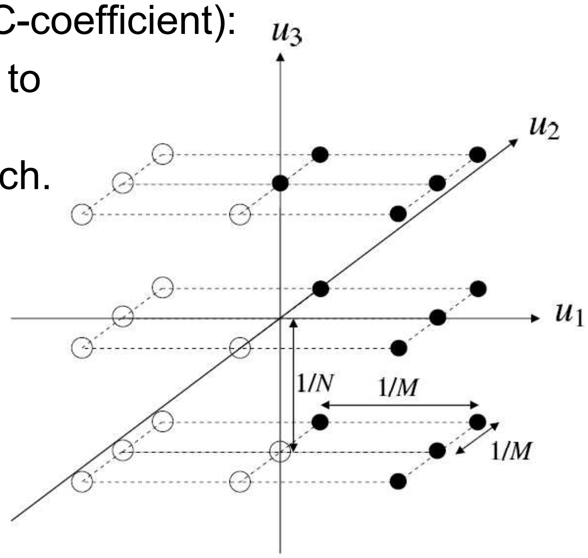
- A local patch in a video is considered as a 3-D volume.
- 3-D STFT can be computed and the phase can be extracted.
- Example:



Original frames Magnitude part Phase part

Spatio-temporal volume LPQ

- For a 3-D descrete domain STFT there are 13 independent low frequency coefficient (excluding the DC-coefficient):
- After 2-bit quantization this would lead to 26-bit presentation, and the histogram would be $2^{26} \approx 6.7e+7$ which is too much.
- PCA is applied to compress the representation.
- The procedure is almost the same as with the 2-D case.
- Data vector is reduced from 26 to 10, which leads to a 1024-bin histogram.



Reference: Päivärinta VJ, Rahtu E & Heikkilä J (2011) Volume local phase quantization for blur-insensitive dynamic texture classification. SCIA 2011, LNCS 6688, pp. 360–369.

Spatio-temporal model

- The covariance between the pixels values in a 3-D volume is assumed to follow the exponential model:

$$\sigma_{ij} = \rho_s^{d_{ij}^s} \rho_t^{d_{ij}^t}$$

where

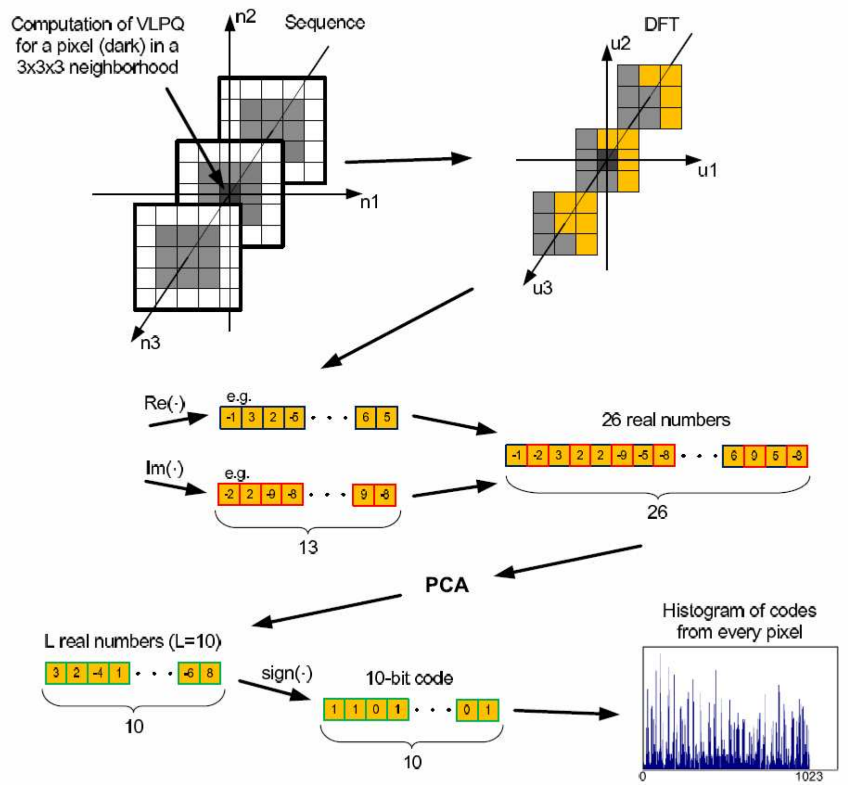
$$d_{ij}^s = \sqrt{\sum_{p=1}^{2} |\mathbf{x}_i(p) - \mathbf{x}_j(p)|^2} \quad \text{and} \quad d_{ij}^t = |\mathbf{x}_i(3) - \mathbf{x}_j(3)|$$

ρ_s and ρ_t are the correlation coefficients of spatially and temporally adjacent pixels, respectively.

- Only 10 most significant eigenvectors are selected for the PCA from

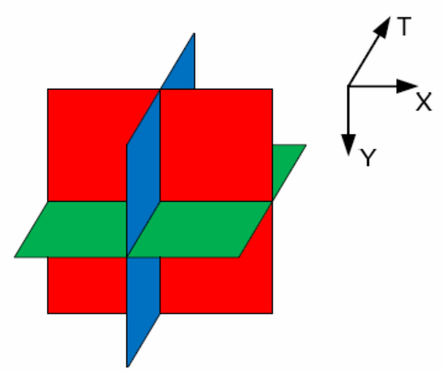
$$\mathbf{C} = \begin{bmatrix} 1 & \sigma_{1,2} & \cdots & \sigma_{1,M^2N} \\ \sigma_{2,1} & 1 & \cdots & \sigma_{2,M^2N} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{M^2N,1} & \sigma_{M^2N,2} & \cdots & 1 \end{bmatrix}$$

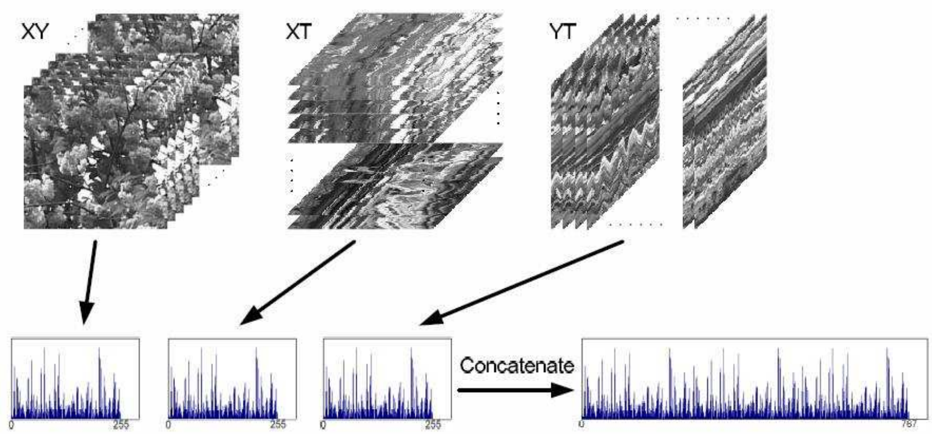
Volume LPQ (VLPQ)



LPQ-TOP

- First proposed by Jiang et al. [FG 2011].
- Similar to LBP-TOP
- LBP has been replaced with LPQ.
- LPQ is computed in three planes XY, XT, and YT.
- Histograms are concatenated.





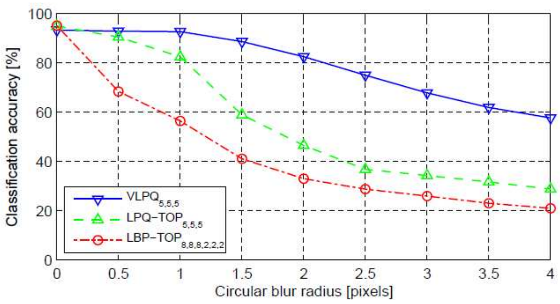
Experiments

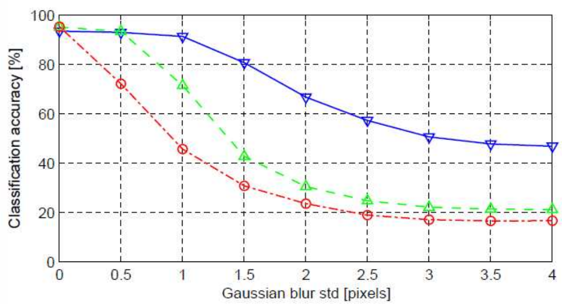
- Experiments are made using the DynTex++ database from

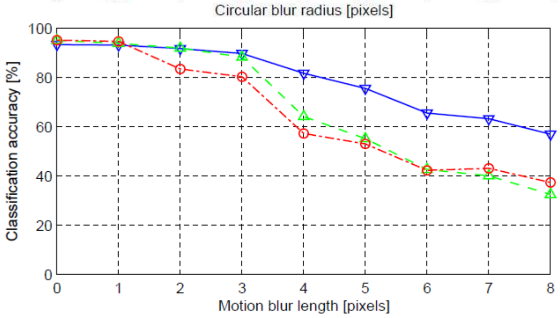
  Ghanem B. & Ahuja N. (2010) Maximum margin distance learning for dynamic exture recognition. In: K. Daniilidis, P. Maragos & N. Paragios (eds.) Computer Vision - ECCV 2010, Lecture Notes in Computer Science, vol. 6312, Springer Berlin / Heidelberg, pp. 223–236.
- 3600 gray scale dynamic textures of size 50×50×50. The textures are divided into 36 classes, each holding 100 videos.
- Examples:

Experimental results

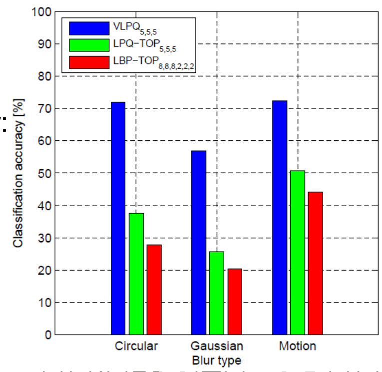
- Correlation coefficients ρ_s = 0.1, ρ_t = 0.1, uniform weighting, window size 5 x 5 x 5.







Spatially and temporally varying blur:



Notice: the maximum margin distance learning technique (Ghanem B. & Ahuja N. ECCV-2010) achieved only 63.7 % accuracy with sharp images.

Implementation and performance

- Matlab implementations of LPQ, RI-LPQ, VLPQ and LPQ-TOP can be downloaded from *http://www.cse.oulu.fi/Downloads/LPQMatlab*
- Execution times for a DynTex++ video sequence (50 x 50 x 50 pixels):

Neighborhood size	VLPQ	LPQ-TOP	LBP-TOP
3	0.13 s	0.28 s	0.15 s
5	0.13 s	0.31 s	0.15 s
7	0.14 s	0.29 s	0.14 s
9	0.14 s	0.29 s	0.13 s
11	0.13 s	0.28 s	0.13 s

Platform: MATLAB R2010a on a 2.4 GHz, 96 GB Sunray server

Papers where LPQ has been used

- Chan C, Kittler J, Poh N, Ahonen T & Pietikäinen M (2009) (Multiscale) local phase quantization histogram discriminant analysis with score normalisation for robust face recognition, In Proc. *IEEE Workshop on Video-Oriented Object and Event Classification,* 633–640.
- Nishiyama M, Hadid A, Takeshima H, Shotton J, Kozakaya T & Yamaguchi O (2011) Facial deblur inference using subspace analysis for recognition of blurred faces. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33(4): 838-845.
- Brahnam S, Nanni L, Shi J-Y & Lumini A (2010) Local phase quantization texture descriptor for protein classification., Proc. *International Conference on Bioinformatics and Computational Biology (Biocomp2010)*, Las Vegas, Nevada, USA, 7 p.
- Nanni L, Lumini A & Brahnam S (2010) Local binary patterns variants as texture descriptors for medical image analysis, *Artificial Intelligence in Medicine* 49(2):117-125.
- Fiche C, Ladret P & Vu NS (2010) Blurred Face Recognition Algorithm Guided by a No-Reference Blur Metric. *Image Processing: Machine Vision Applications III*, 9 p.
- Jiang B, Valstar MF & Pantic M (2011) Action unit detection using sparse appearance descriptors in space-time video volumes. Proc. *9th IEEE Conference on Automatic Face and Gesture Recognition (FG 2011),* Santa Barbara, CA, 314-321.
- Yang S & Bhanu B (2011) Facial Expression Recognition Using Emotion Avatar Image. Proc. *Workshop on Facial Expression Recognition and Analysis Challenge FERA2011,* Santa Barbara, CA, 866-871.
- Dhall A., Asthana A., Goecke R., and Gedeon T. (2011) Emotion Recognition Using PHOG and LPQ features. Proc. *Workshop on Facial Expression Recognition and Analysis Challenge FERA2011*, Santa Barbara (CA), 878-883.

Conclusions

- Many previous works exist where phase information has been utilized in image analysis.
- Our contribution is the framework for constructing blur insensitive texture descriptors that are both robust and computationally efficient.
- LPQ and its variants can be used for characterizing blurred still images and videos.
- Good performance is also achieved with non-blurred data.
- STFT with uniform weighting seems to be a good choice for computing the phase.
- Filters can be truncated without loss of important information.
- Decorrelation improves the accuracy when blur is isotropic.
- LPQ has been already used by many researchers in fields of medical image analysis and facial image analysis.

Part 4: Example applications

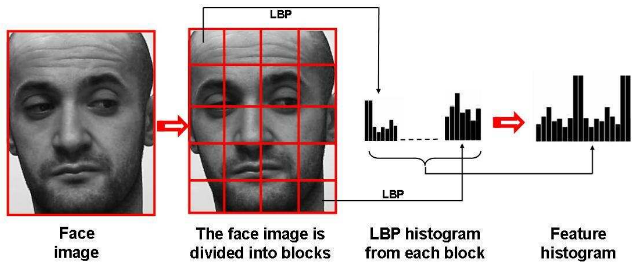
Matti Pietikäinen

Infotech Oulu MACHINE VISION GROUP UNIVERSITY of OULU OULUN YLIOPISTO

Face analysis using local binary patterns

- Face recognition is one of the major challenges in computer vision

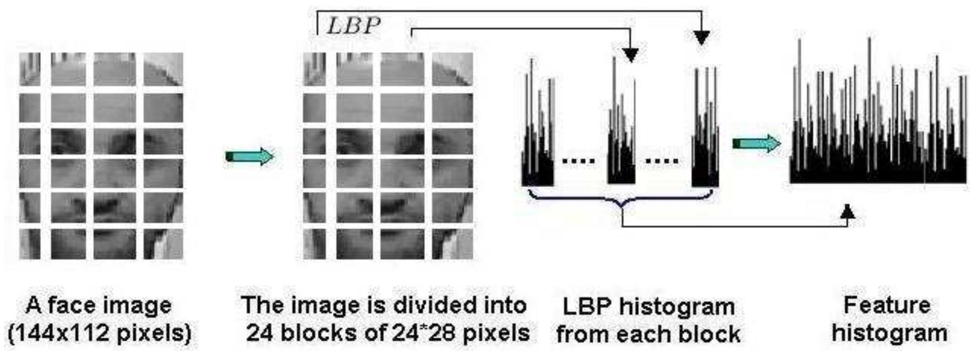
- Ahonen et al. proposed (ECCV 2004, PAMI 2006) a face descriptor based on LBPs
- This method has already been adopted by many leading scientists and groups
- Computationally very simple, excellent results in face recognition and authentication, face detection, facial expression recognition, gender classification



Face description with LBP

Ahonen T, Hadid A & Pietikäinen M (2006) Face description with local binary patterns: application to face recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence 28(12):2037-2041. (an early version published at ECCV 2004)

A facial description for face recognition:



Weighting the regions

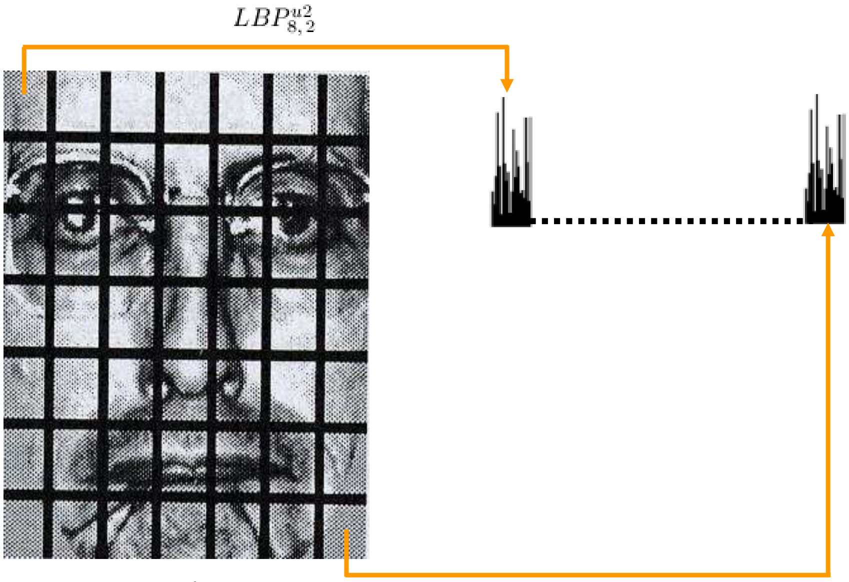
Block size

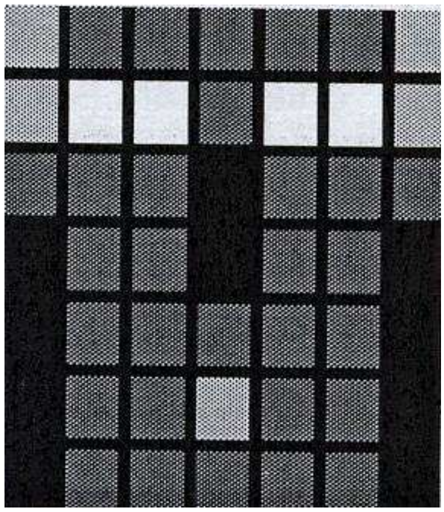
18 * 21

Metrics

χ^2 (Chi-square)

Weighting



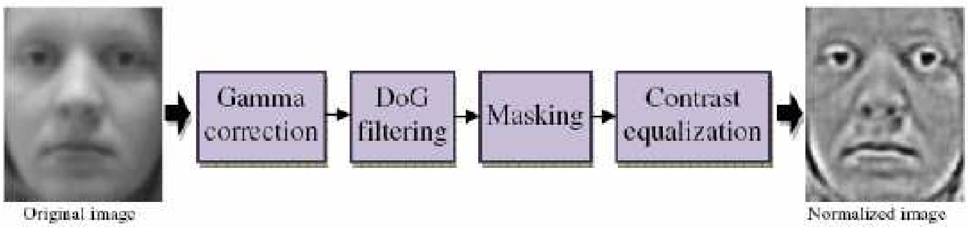


130 * 150

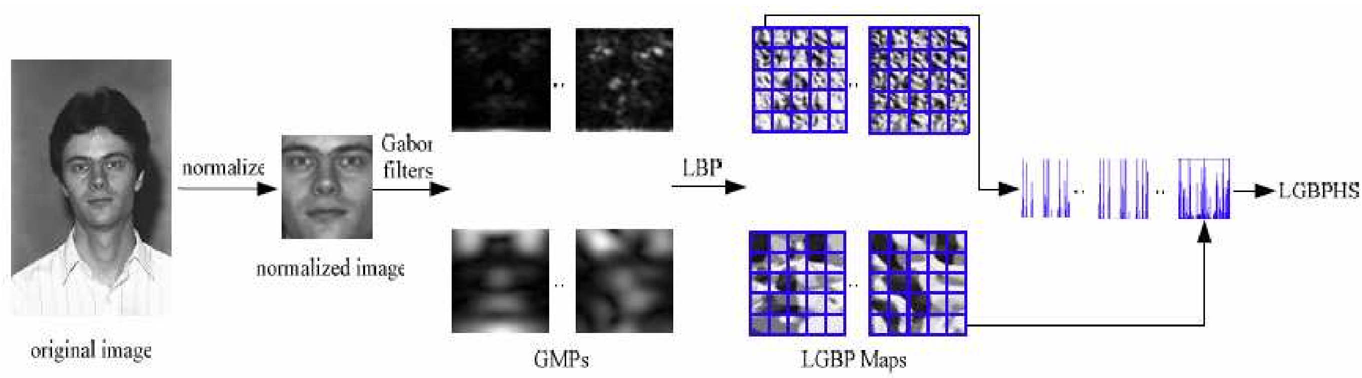
Feature vector length → 2891

Improving the robustness of LBP-based face recognition

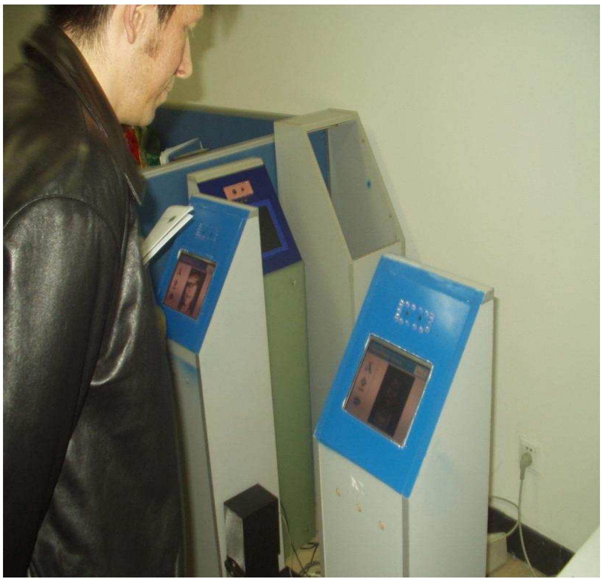
Illumination normalization by preprocessing] prior to LBP/LPQ feature extraction [Tan & Triggs, AMFG 2007]



Local Gabor Binary Pattern Histogram Sequence [Zhang et al., ICCV 2005]



Illumination invariance using LBP with NIR imaging S.Z. Li et al. [IEEE PAMI, 2007]



Strong illumination | Weak illumination | Dark illumination

Case: Trusted biometrics under spoofing attacks (TABULA RASA) 2010-2014 (http://www.tabularasa-euproject.org/)

- The project will address some of the issues of direct (spoofing) attacks to trusted biometric systems. This is an issue that needs to be addressed urgently because it has recently been shown that conventional biometric techniques, such as fingerprints and face, are vulnerable to direct (spoof) attacks.

- Coordinated by IDIAP, Switzerland

- We will focus on face and gait recognition

SEVENTH FRAMEWORK PROGRAMME

TABULA RASA

Example of 2D face spoofing attack

Figure 1. Example of a 2D face spoofing attack using a photograph.

Figure 2. Example of images captured from real faces (upper row) and from printed photos (lower row). The appearance similarity illustrates the difficulty of spoofing detection from printed photos.

LBP is very powerful, discriminating printing artifacts and differences in light reflection

- outperformed results of Tan et al. [ECCV 2010], and LPQ and Gabor features

Automatic landscape mode detection

Huttunen S, Rahtu E, Heikkilä J, Kunttu I & Gren J (2011) Real-time detection of landscape scenes. Proc. Scandinavian Conference on Image Analysis (SCIA 2011), LNCS, 6688:338-347.

- The aim was to develop and implement an algorithm that automatically classifies images to *landscape* and *non-landscape* categories
  - The analysis is solely based on the visual content of images.
- The main criterion is to find an accurate but still computationally light solution capable of real-time operation.

Landscape vs. non-landscape

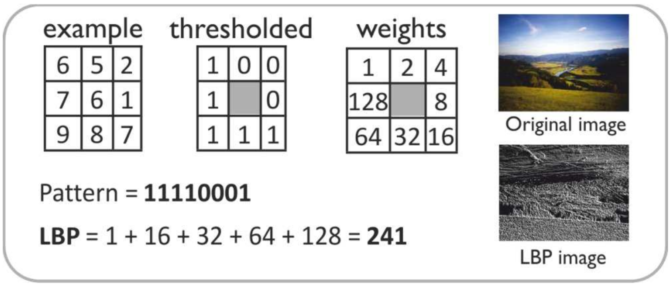
- Definition of *landscape* and *non-landscape* images is not straightforward
  - If there are no distinct and easily separable objects present in a natural scene, the image is classified as landscape
  - The non-landscape branch consists of indoor scenes and other images containing man-made objects at relatively close distance

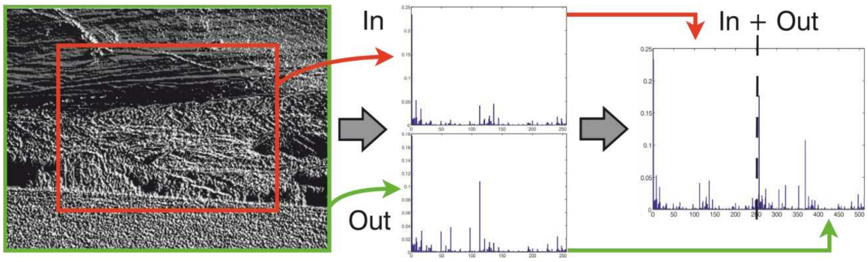
Data set

- The images used for training and testing were downloaded from the PASCAL Visual Object Classes (VOC2007) database and the Flickr site
  - All the images were manually labeled and resized to QVGA (320x240).
- Training: ***1115 landscape*** images and
  2617 non-landscape images
- Testing: ***912*** and ***2140***, respectively

The approach

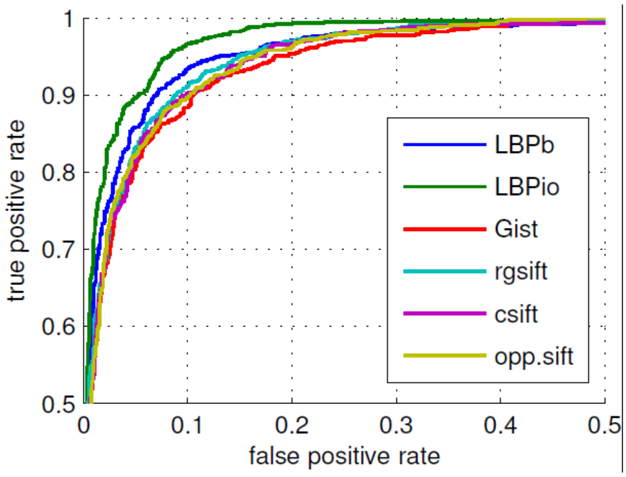
- Simple global image representation based on local binary pattern (LBP) histograms is used
- Two variants:
  - Basic LBP
  - LBP In+Out
- SVM classifier

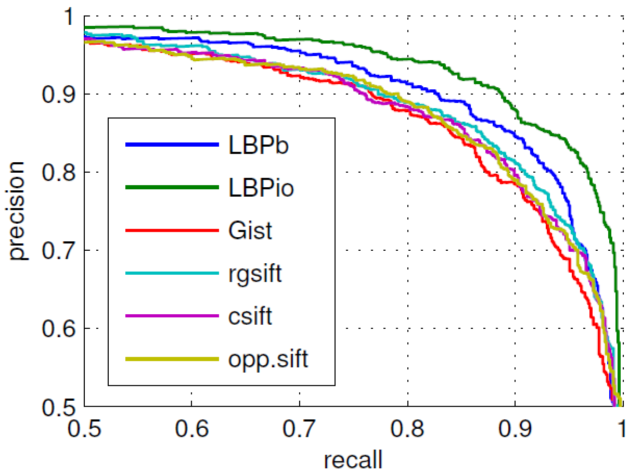




Feature extraction → **Histogram computation** → **SVM classifier training**

Classification results





Classification examples

Landscape as landscape (TP)

Non-landscape as non-landscape (TN)

Non-landscape as landscape (FP)

Landscape as non-landscape (FN)

Summary of the results

		Classification				Execution time (s)	
		AUC	AP	TPR	FPR	Descriptor	Total
Global	LBP_{io}	0.982	0.958	0.882	0.040	0.001	0.005
	LBP_b	0.972	0.939	0.862	0.055	0.001	0.003
	GIST	0.963	0.924	0.809	0.050	NA	>0.029
Local descriptors + BOW	rgsift	0.969	0.934	0.814	0.045	0.340	2.699
	csift	0.966	0.926	0.825	0.052	0.350	2.712
	opponentsift	0.966	0.922	0.828	0.052	0.340	2.694
	rgbsift	0.960	0.918	0.804	0.047	0.330	2.744
	hsvsift	0.959	0.915	0.804	0.050	0.340	2.494
	sift	0.956	0.901	0.806	0.059	0.120	0.595
	huesift	0.954	0.902	0.791	0.067	0.290	1.046
	colormomentinvariants	0.926	0.857	0.737	0.067	1.410	1.444
	transformedcolorhistogram	0.924	0.851	0.692	0.061	0.100	0.147
	opponenthistogram	0.909	0.825	0.689	0.079	0.090	0.140
	rgbhistogram	0.903	0.805	0.683	0.087	0.070	0.118
	colormoments	0.897	0.811	0.697	0.084	1.340	1.376
	huehistogram	0.863	0.717	0.525	0.071	0.180	0.223
	nrghistogram	0.861	0.727	0.601	0.102	0.080	0.119

Real-time implementation

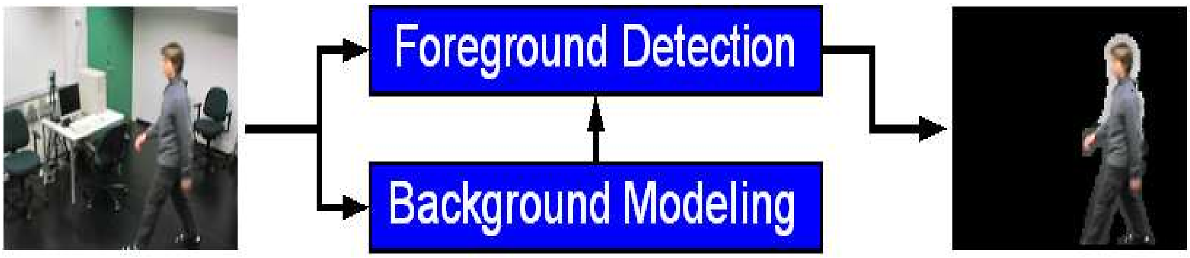
- The current real-time implementation coded in C relies on the basic LBP_b
- Performance analysis
  - Windows PC with Visual Studio 2010 Profiler
    - The total execution time for one frame was about **3 ms**
  - Nokia N900 with FCam
    - About **30 ms**

Demo videos

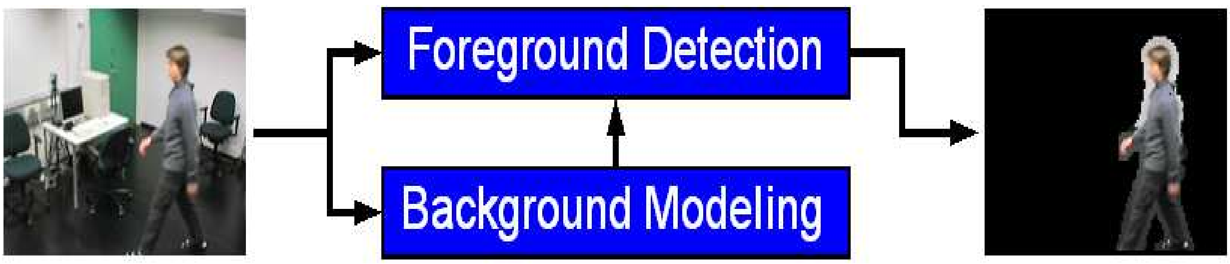
Reference: Huttunen S, Rahtu E, Kunttu I, Gren J & Heikkilä J (2011) Real-time detection of landscape scenes. Proc. Scandinavian Conference on Image Analysis (SCIA 2011), LNCS, 6688:338-347.

Modeling the background and detecting moving objects

Heikkilä M & Pietikäinen M (2006) A texture-based method for modeling the background and detecting moving objects. IEEE Transactions on Pattern Analysis and Machine Intelligence 28(4):657-662. (an early version published at BMVC 2004)



Roughly speaking, the background subtraction can be seen as a two-stage process as illustrated below.



Background modeling

The goal is to construct and maintain a statistical representation of the scene that the camera sees.

Foreground Detection

The comparison of the input frame with the current background model. The areas of the input frame that do not fit to the background model are considered as foreground.

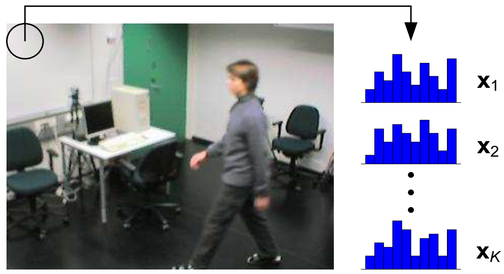
…Overview of the approach…

We use an LBP histogram computed over a circular region around the pixel as the feature vector.

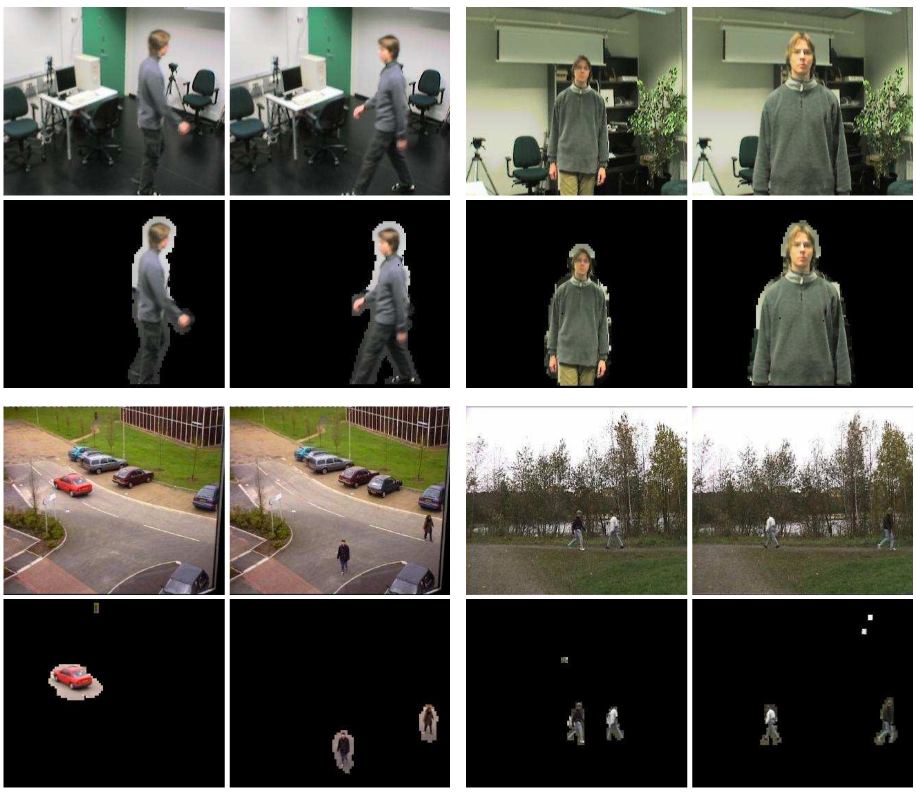
The history of each pixel over time is modeled as a group of K weighted LBP histograms: $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_K\}$.

The background model is updated with the information of each new video frame, which makes the algorithm adaptive.

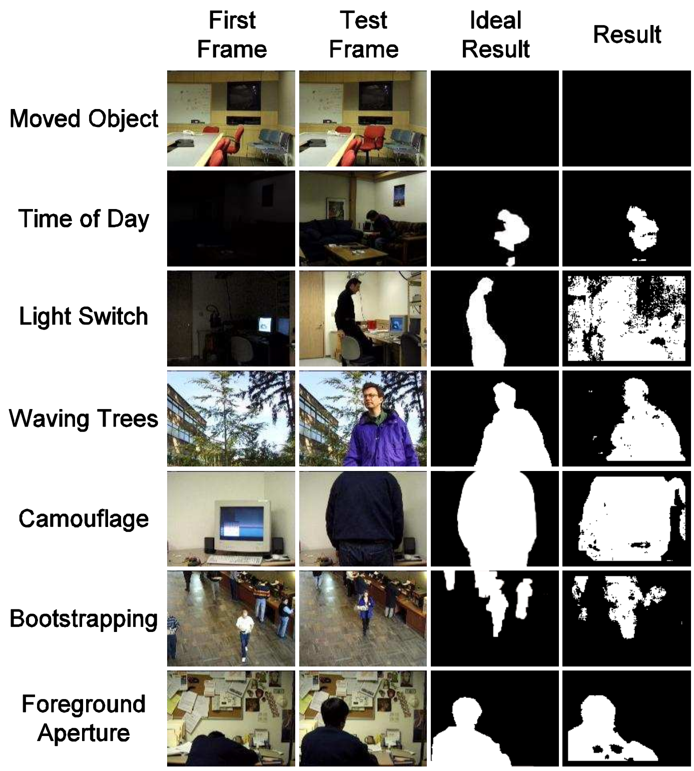
The update procedure is identical for each pixel.



Examples of detection results



Detection results for images of Toyama et al. (ICCV 1999)

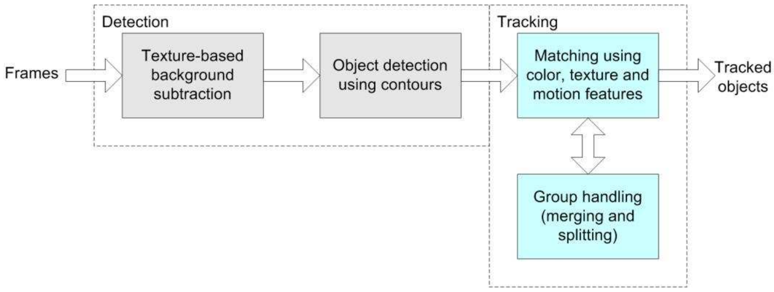


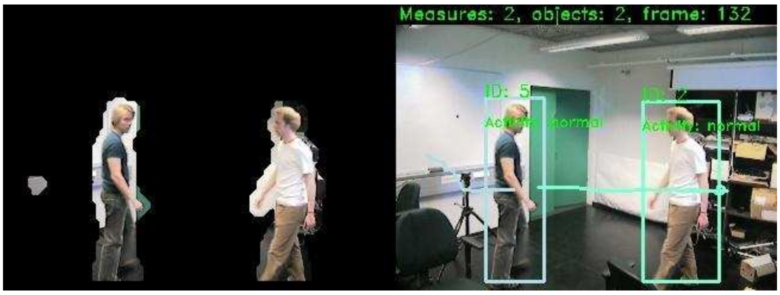
Demo for detection of moving objects

Infotech Oulu
MACHINE VISION GROUP
UNIVERSITY of OULU
OULUN YLIOPISTO

LBP in multi-object tracking

Takala V & Pietikäinen M (2007) Multi-object tracking using color, texture, and motion. Proc. Seventh IEEE International Workshop on Visual Surveillance (VS 2007), Minneapolis, USA, 7 p.

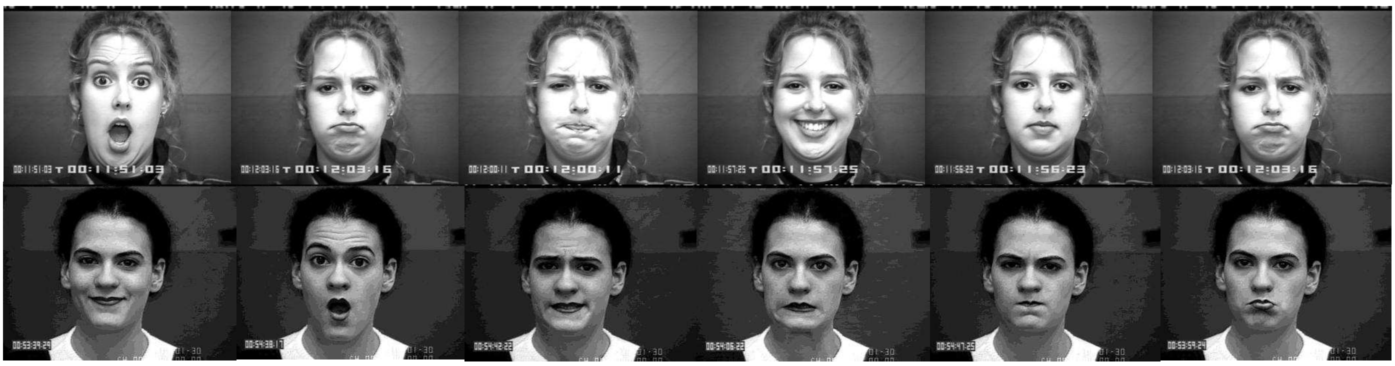




Facial expression recognition from videos

Zhao G & Pietikäinen M (2007) Dynamic texture recognition using local binary patterns with an application to facial expressions. IEEE Transactions on Pattern Analysis and Machine Intelligence 29(6):915-928.

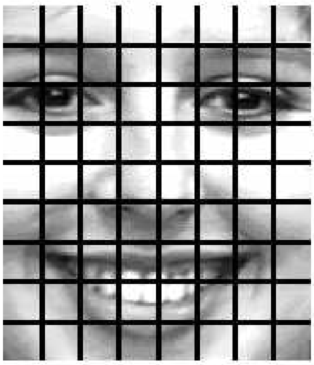
- Determine the emotional state of the face
  - Regardless of the identity of the face



Facial Expression Recognition

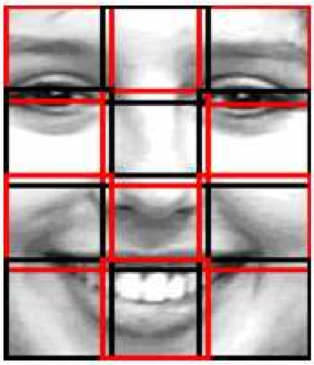
- Mug Shot
  - [Feng, 2005][Shan, 2005]
  - [Bartlett, 2003][Littlewort,2004]
- Dynamic Information
  - Action Units
    - [Tian, 2001][Lien, 1998]
    - [Bartlett,1999][Donato,1999]
    - [Cohn,1999]
  - Prototypic Emotional Expressions
    - [Cohen,2003]
    - [Yeasin, 2004]
    - [Aleksic,2005]

Psychological studies [Bassili 1979], have demonstrated that humans do a better job in recognizing expressions from dynamic images as opposed to the mug shot.



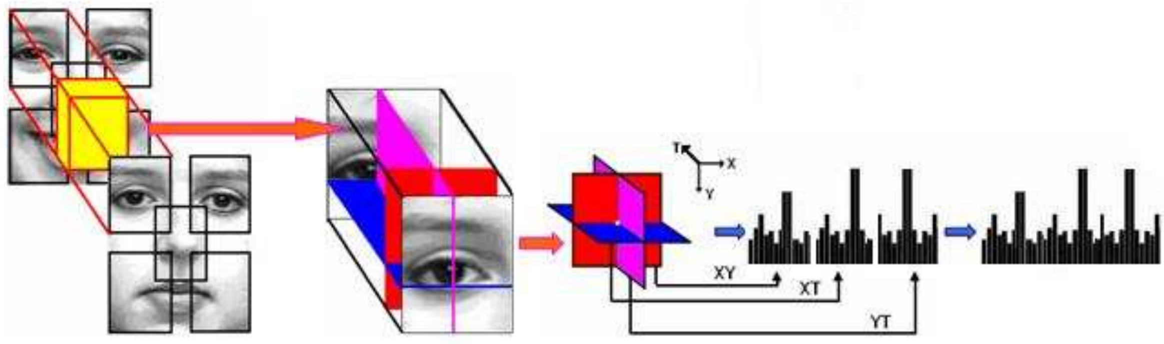
(a)

(a) Non-overlapping blocks(9 x 8)



(b)

(b) Overlapping blocks (4 x 3, overlap size = 10)



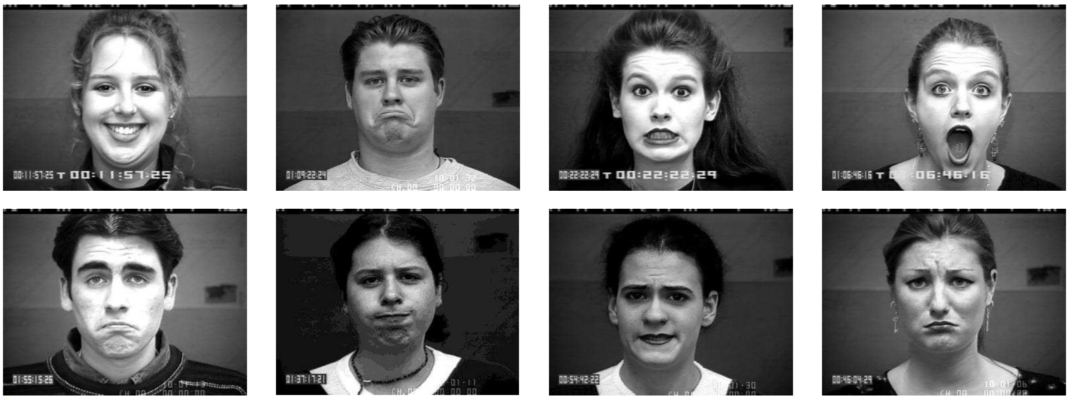
(a) (b) (c)

(a) Block volumes (b) LBP features from three orthogonal planes (c) Concatenated features for one block volume with the appearance and motion

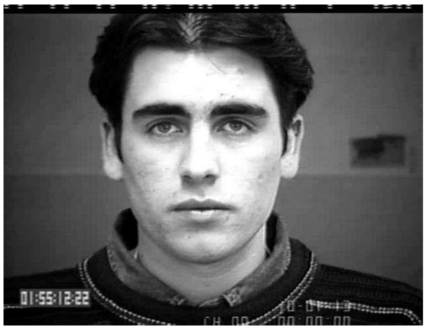
Database

Cohn-Kanade database :

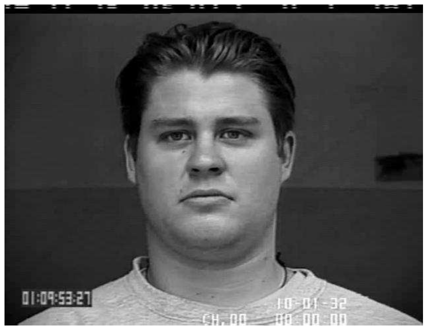
- 97 subjects
- 374 sequences
- Age from 18 to 30 years
- Sixty-five percent were female, 15 percent were African-American, and three percent were Asian or Latino.



Happiness



Anger



Disgust

Sadness

Fear

Surprise

Infotech Oulu
MACHINE VISION GROUP
UNIVERSITY of OULU
OULUN YLIOPISTO

Comparison with different approaches

	People Num	Sequence Num	Class Num	Dynamic	Measure	Recognition Rate (%)
[Shan,2005]	96	320	7(6)	N	10 fold	88.4(92.1)
[Bartlett, 2003]	90	313	7	N	10 fold	86.9
[Littlewort, 2004]	90	313	7	N	leave-one-subject-out	93.8
[Tian, 2004]	97	375	6	N	-------	93.8
[Yeasin, 2004]	97	------	6	Y	five fold	90.9
[Cohen, 2003]	90	284	6	Y	-------	93.66
Ours	97	374	6	Y	two fold	**95.19**
Ours	97	374	6	Y	10 fold	**96.26**

Demo for facial expression recognition

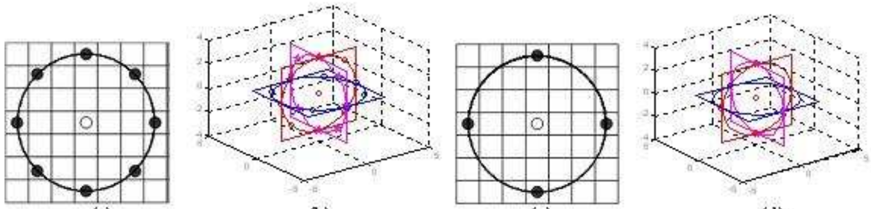
- **Low resolution**
- **No eye detection**
- **Translation, in-plane and out-of-plane rotation, scale**
- **Illumination change**
- **Robust with respect to errors in face alignment**

Principal appearance and motion from boosted spatiotemporal descriptors

Zhao G & Pietikäinen M (2009) Boosted multi-resolution spatiotemporal descriptors for facial expression recognition. Pattern Recognition Letters 30(12):1117-1127.

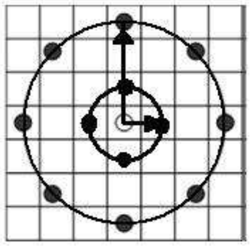
Multiresolution features=>Learning for pairs=>Slice selection

- 1) Use of different number of neighboring points when computing the features in XY, XT and YT slices



(a) (b) (c) (d)

- 2) Use of different radii which can catch the occurrences in different space and time scales



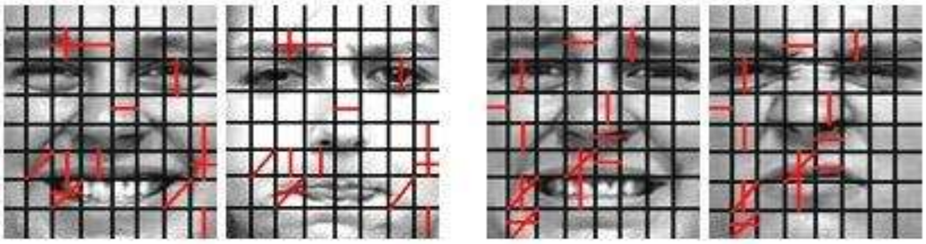
- 3) Use of blocks of different sizes to have global and local statistical features

The first two resolutions focus on the
❖ pixel level in feature computation, providing different local spatiotemporal information

the third one focuses on the
❖ block or volume level, giving more global information in space and time dimensions.

Infotech Oulu
MACHINE VISION GROUP
UNIVERSITY of OULU
OULUN YLIOPISTO

Selected 15 most discriminative slices



Happiness vs. Anger　　　Fear vs. Disgust

Example images in different illuminations

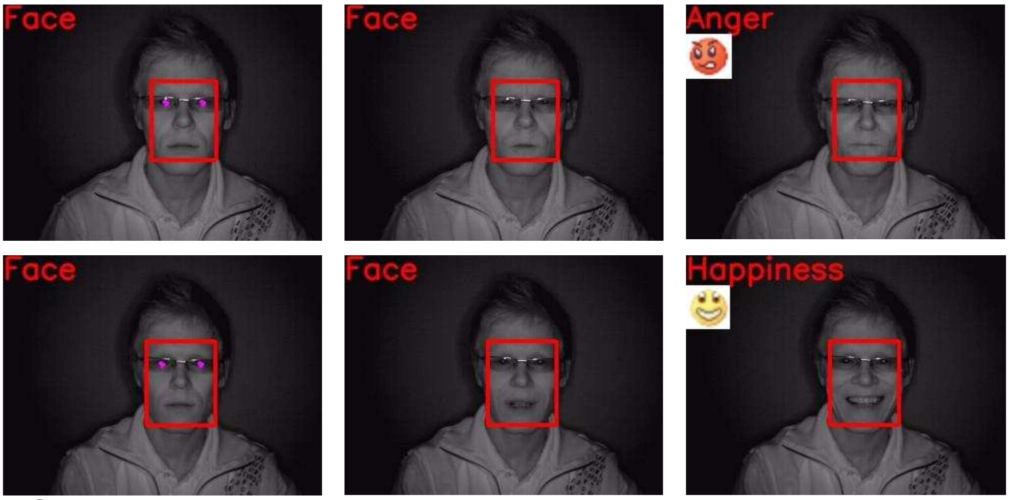
Visible light (VL) : 0.38-0.75 μm
Near Infrared (NIR) : 0.7μm-1.1μm

Strong illumination　　Weak illumination　　Dark illumination

Taini M, Zhao G, Li SZ & Pietikäinen M (2008) Facial expression recognition from near-infrared video sequences. Proc. International Conference on Pattern Recognition (ICPR), 4 p.

On-line facial expression recognition from NIR videos

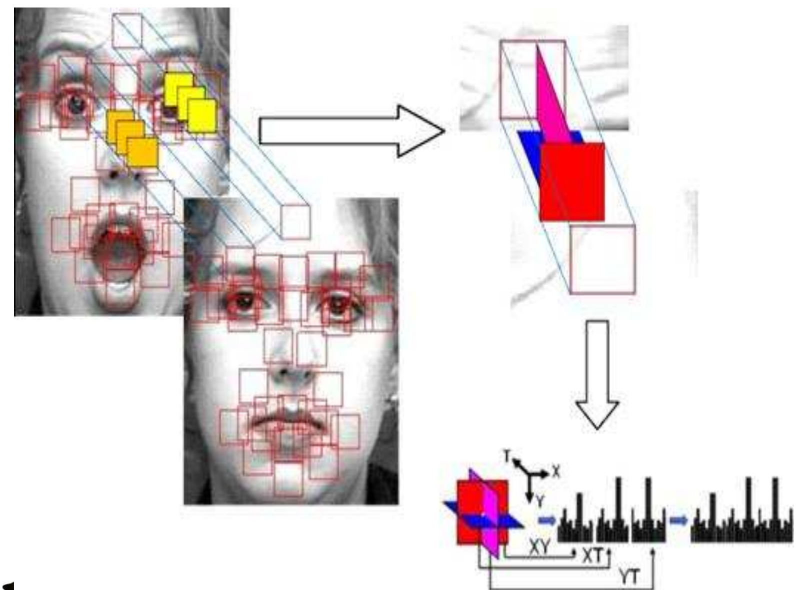
- NIR web camera allows expression recognition in near darkness.
- Image resolution 320 × 240 pixels.
- 15 frames used for recognition.
- Distance between the camera and subject around one meter.

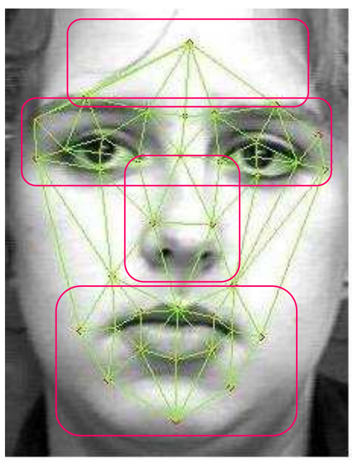


Start sequences Middle sequences End sequences

Component-based approaches [Huang et al.,2010-2011]

Boosted spatiotemporal LBP-TOP features are extracted from areas centered at fiducial points (detected by ASM) or larger areas
- more robust to changes of pose, occlusions
- can be used for analyzing action units [Jiang et al, FG 2011]





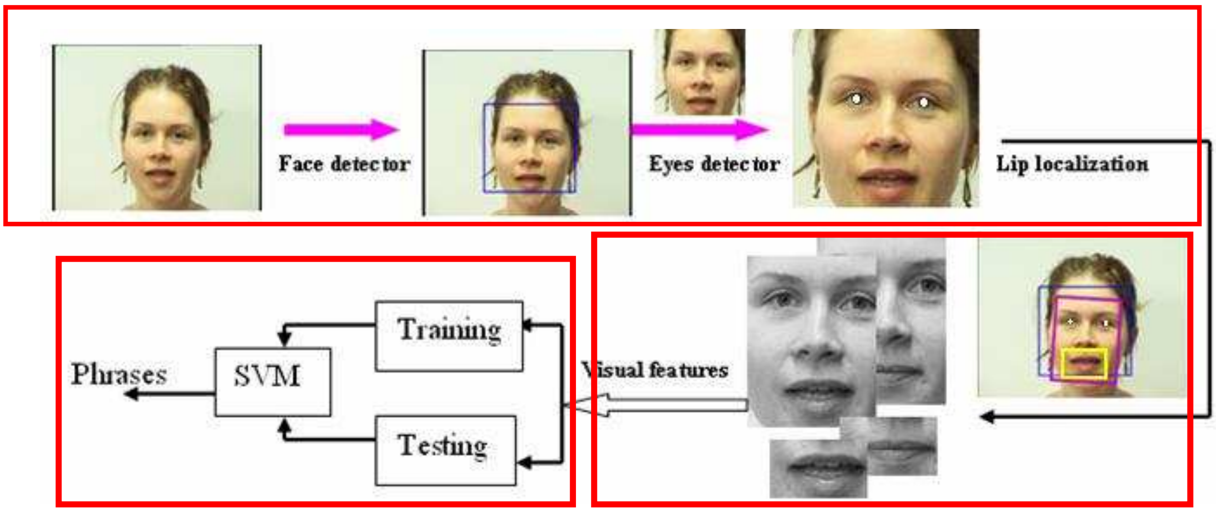
Visual speech recognition

Zhao G, Barnard M & Pietikäinen M (2009). Lipreading with local spatiotemporal descriptors. IEEE Transactions on Multimedia 11(7):1254-1265.

- Visual speech information plays an important role in speech recognition under noisy conditions or for listeners with hearing impairment.
- A human listener can use visual cues, such as lip and tongue movements, to enhance the level of speech understanding.
- The process of using visual modality is often referred to as lipreading which is to make sense of what someone is saying by watching the movement of his lips.

McGurk effect [McGurk and MacDonald 1976] demonstrates that inconsistency between audio and visual information can result in perceptual confusion.

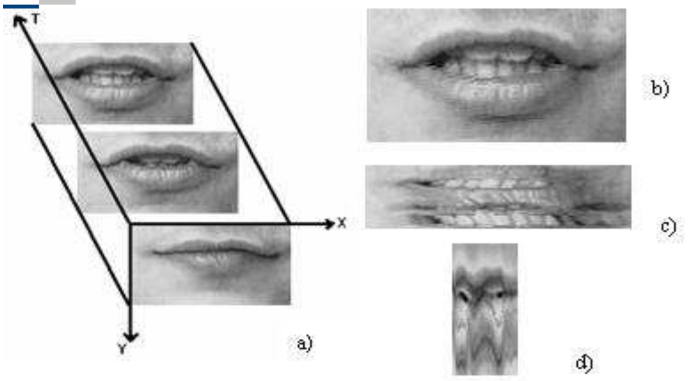
System overview



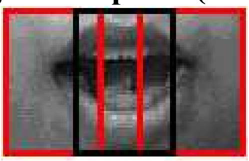
Our system consists of three stages.

- First stage: face and eye detectors, and the localization of mouth.
- Second stage: extracts the visual features.
- Last stage: recognize the input utterance.

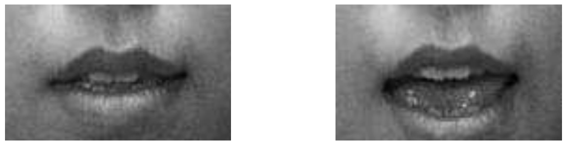
Local spatiotemporal descriptors for visual information



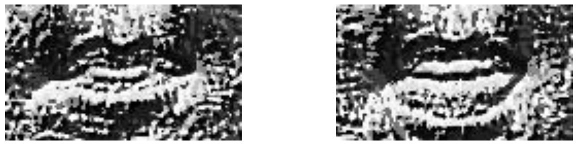
(a) Volume of utterance sequence
(b) Image in XY plane (147x81)
(c) Image in XT plane (147x38) in y =40
(d) Image in TY plane (38x81) in x = 70



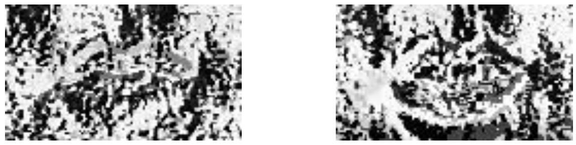
Overlapping blocks (1 x 3, overlap size = 10).



Mouth region images

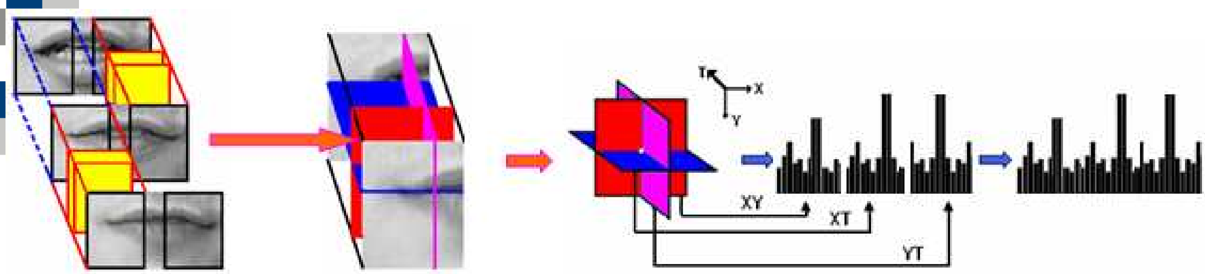


LBP-XY images



LBP-XT images

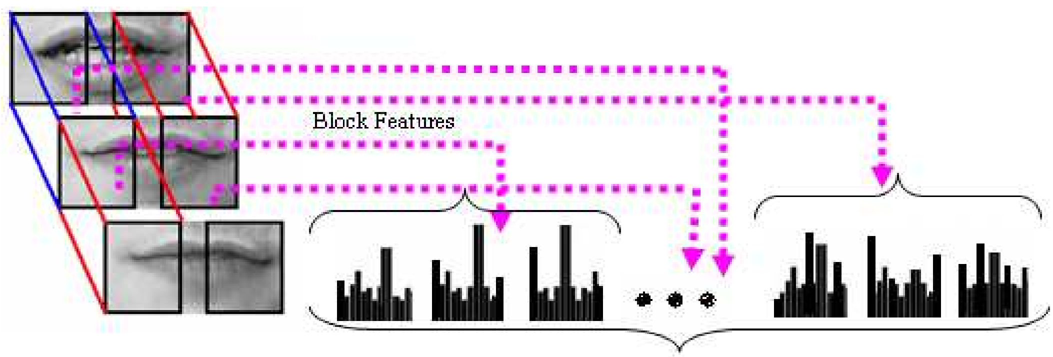
LBP-YT images



(a) (b) (c)

(a) Block volumes (b) LBP features from three orthogonal planes (c) Concatenated features for one block volume with the appearance and motion

Features in each block volume.



Mouth movement features from the whole sequence

Mouth movement representation.

Experiments

Database:

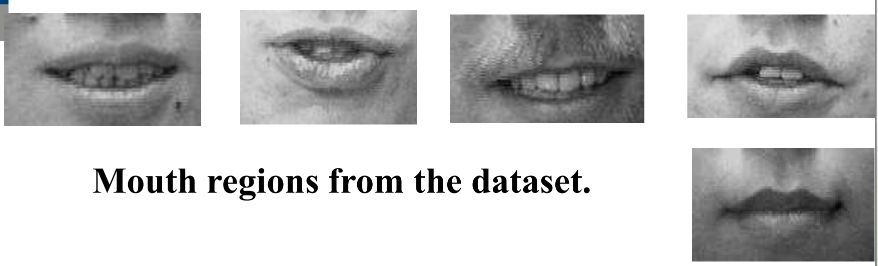
Our own visual speech database: OuluVS Database

20 persons; each uttering ten everyday's greetings one to five times.

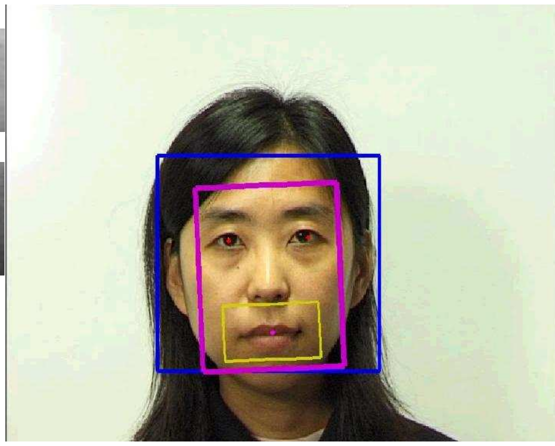
Totally, 817 sequences from 20 speakers were used in the experiments.

C1	"Excuse me"	C6	"See you"
C2	"Good bye"	C7	"I am sorry"
C3	"Hello"	C8	"Thank you"
C4	"How are you"	C9	"Have a good time"
C5	"Nice to meet you"	C10	"You are welcome"

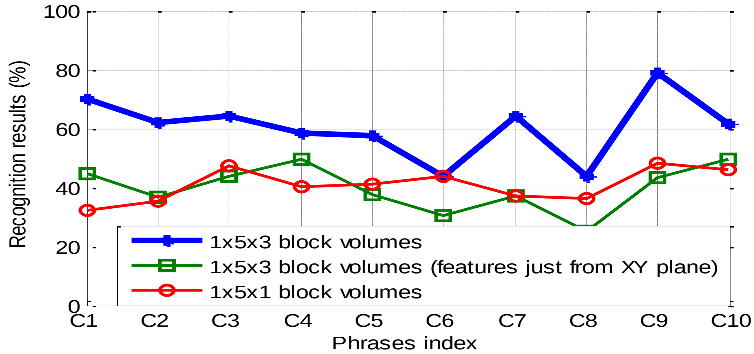
Experimental results - OuluVS database



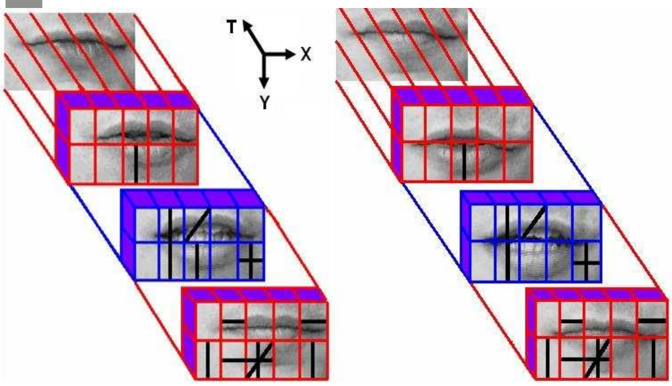
Mouth regions from the dataset.



Speaker-independent:



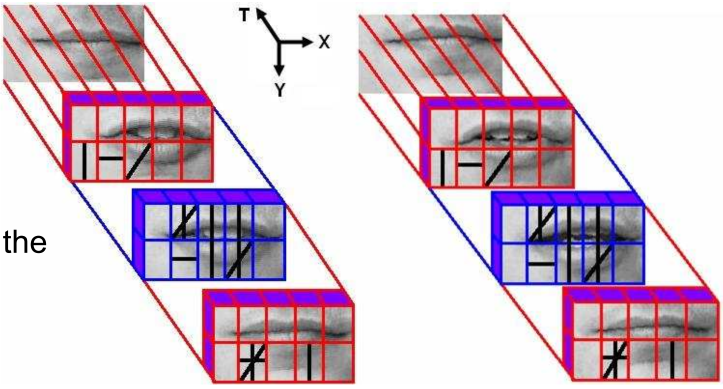
Selecting 15 most discriminative features



These phrases were most difficult to recognize because they are quite similar in the latter part containing the same word "you". The selected slices are mainly in the first and second part of the phrase.

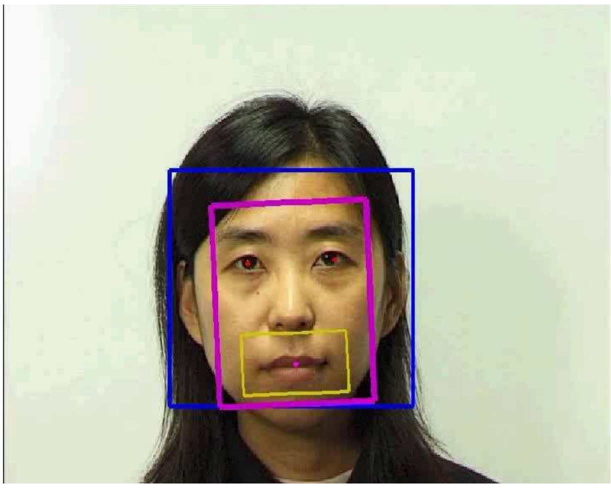
Selected 15 slices for phrases "See you" and "Thank you".

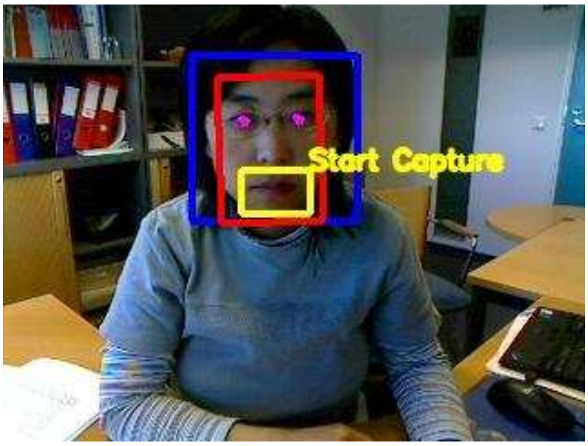
The phrases "excuse me" and "I am sorry"are different throughout the whole utterance, and the selected features also come from the whole pronunciation.



Selected 15 slices for phrases "Excuse me" and "I am sorry".

Demo for visual speech recognition





LBP-TOP with video normalization [Zhou et al., CVPR 2011]

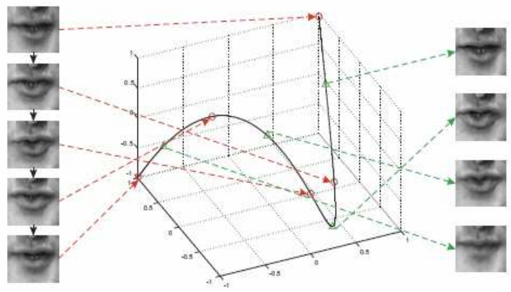


Figure 1. An example of a speech video (left) being projected onto a low-dimensional curve (middle) from which the images (right) are synthesized.

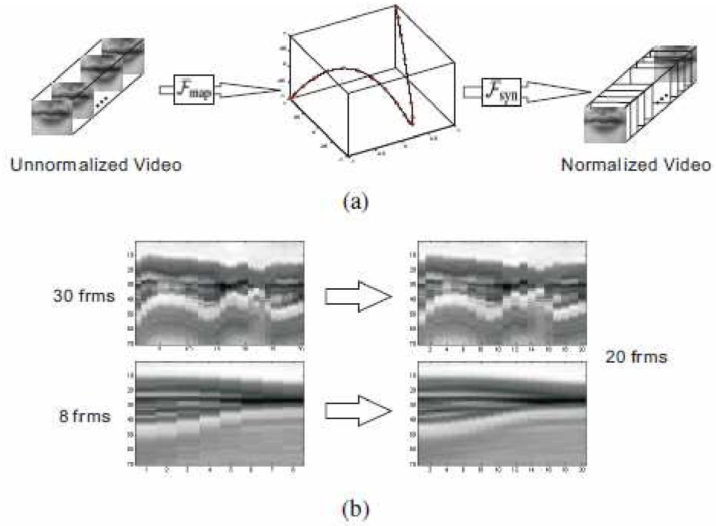
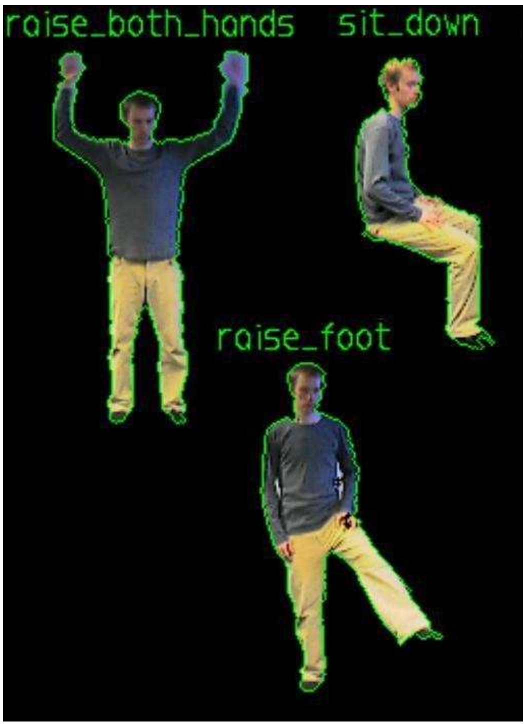


Figure 5. Video normalization: (a) illustrates the normalization process that maps the original video onto a curve and sample a novel video along the curve. (b) shows the temporal patterns of two videos with 8 and 30 frames and their counterparts when both are normalized to be 20-frame long.

With normalization nearly 20% improvement in speaker independent recognition is obtained

Activity recognition

Kellokumpu V, Zhao G & Pietikäinen M (2009) Recognition of human actions using texture. Machine Vision and Applications (available online).

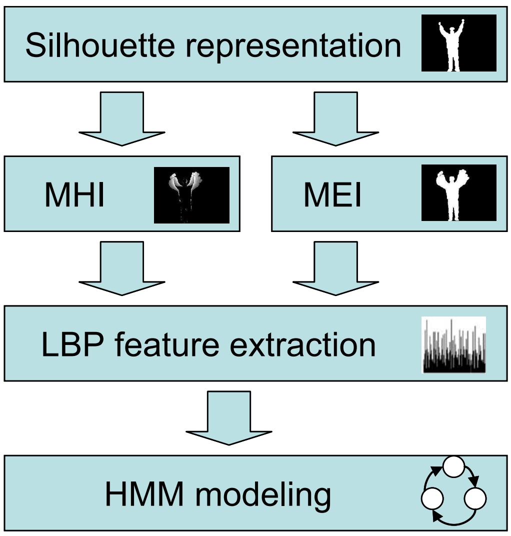


Texture based description of movements

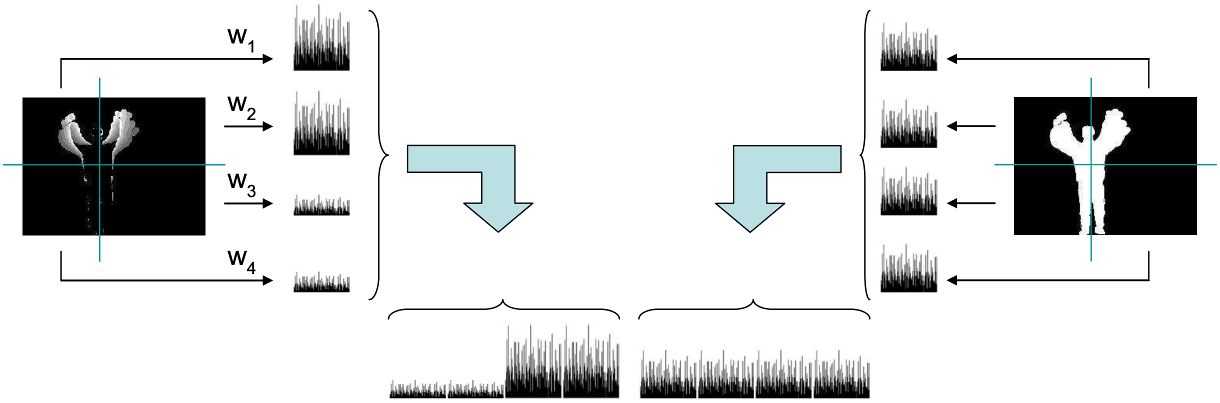
- We want to represent human movement with it's local properties
  > Texture
- But texture in an image can be anything? (clothing, scene background)
  > Need preprocessing for movement representation
  > We use temporal templates to capture the dynamics
- We propose to extract texture features from temporal templates to obtain a short term motion description of human movement.

Kellokumpu V, Zhao G & Pietikäinen M (2008) Texture based description of movements for activity analysis. Proc. International Conference on Computer Vision Theory and Applications (VISAPP), 1:206-213.

Overview of the approach

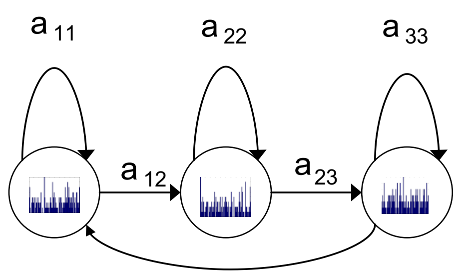


Features



Hidden Markov Models (HMM)

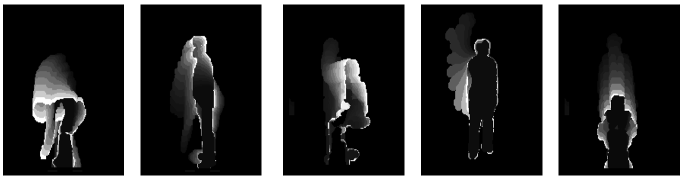
- Model is defined with:
  - Set of observation histograms H
  - Transition matrix A
  - State priors
- Observation probability is taken as intersection of the observation and model histograms:



$$P(h_{obs} \mid s_t = q_i) = \sum \min(h_{obs}, h_i)$$

Experiments

- Experiments on two databases:
  - Database 1:
    - 15 activities performed by 5 persons



  - Database 2 - Weizmann database:
    - 10 Activities performed by 9 persons
    - Walkig, running, jumping, skipping etc.

Experiments – HMM classification

- Database 1 – 15 activities by 5 people
- $LBP_{8,2}$

MHI	99%
MEI	90%
MHI + MEI	100%

- Weizmann database – 10 activities by 9 people
- $LBP_{4,1}$

Ref.	Act.	Seq.	Res.
Our method	**10**	**90**	**97,8%**
Wang and Suter 2007	10	90	**97,8%**
Boiman and Irani 2006	9	81	97,5%
Niebles et al 2007	9	83	72,8%
Ali et al. 2007	9	81	92,6%
Scovanner et al. 2007	10	92	82,6%

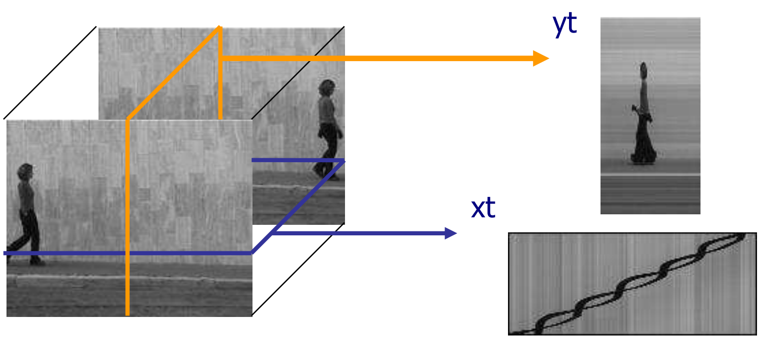
Activity recognition using dynamic textures

- Instead of using a method like MHI to incorporate time into the description, the dynamic texture features capture the dynamics straight from image data.
- When image data is used, accurate segmentation of the silhouette is not needed
  - Instead a bounding box of a person is sufficient!!

Kellokumpu V, Zhao G & Pietikäinen M (2008) Human activity recognition using a dynamic texture based method. Proc. British Machine Vision Conference (BMVC), 10 p.

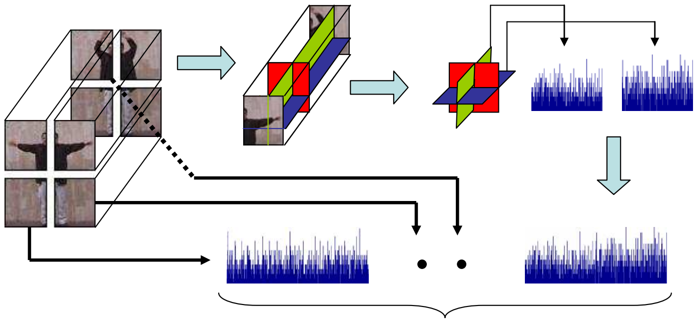
Dynamic textures for action recognition

- Illustration of xyt-volume of a person walking



Dynamic textures for action recognition

- Formation of the feature histogram for an *xyt* volume of short duration



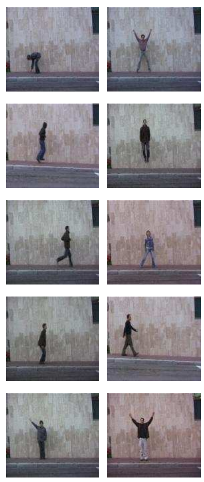
Feature histogram of a bounding volume

- HMM is used for sequential modeling

Action classification results – Weizmann dataset

- Classification accuracy 95,6% using <u>image data</u>

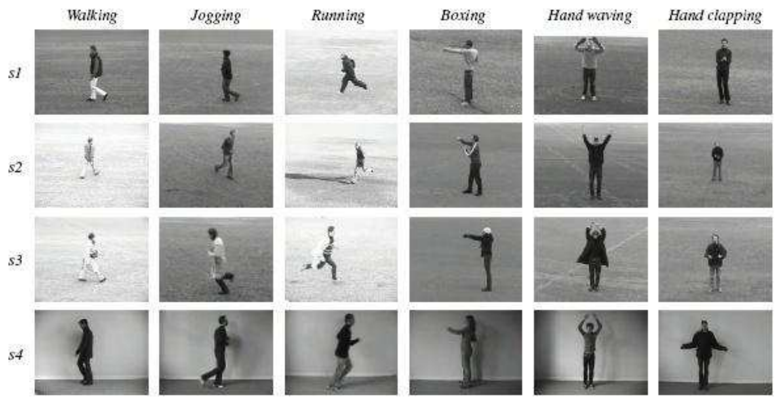
	Bend	Jack	Jump	Pjump	Run	Side	Skip	Walk	Wave1	Wave2
Bend	1.00									
Jack		1.00								
Jump			.78			.11	.11			
Pjump				1.00						
Run					1.00					
Side						1.00				
Skip					.22		.78			
Walk								1.00		
Wave1									1.00	
Wave2										1.00



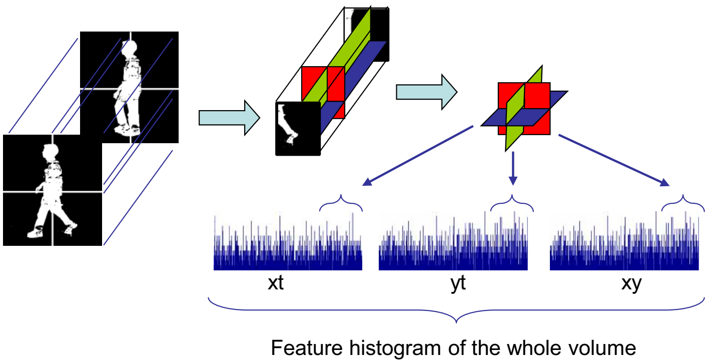
Action classification results - KTH

- Classification accuracy 93,8% using image data

	Box	Clap	Wave	Jog	Run	Walk
Box	.967	.033				
Clap	.003	.987	.01			
Wave	.003	.020	.977			
Jog				.860	,108	.032
Run				.145	.855	
Walk				.020		.980



Dynamic textures for gait recognition



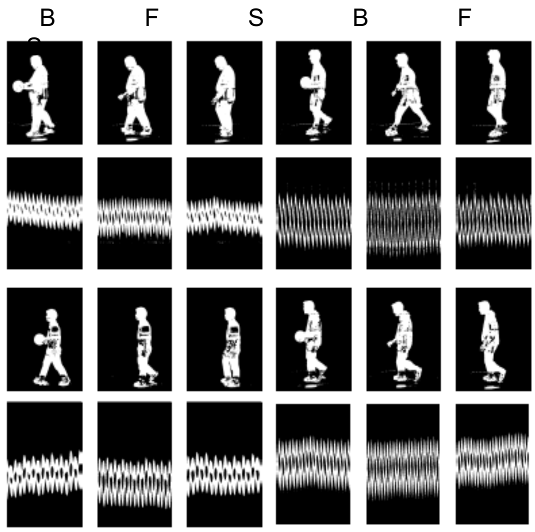
$$Similarity = \sum \min(h_i, h_j)$$

Kellokumpu V, Zhao G & Pietikäinen M (2009) Dynamic texture based gait recognition. Proc. International Conference on Biometrics (ICB), 1000-1009.

Experiments - CMU gait database

CMU database

- 25 subjects
- 4 different conditions (ball, slow, fast, ~~incline~~)



Experiments - Gait recognition results

	S/B	B/S	F/B	B/F	S/F	F/S
CMU [4]	**92 %**	-	-	-	76 %	-
UMD [5]	48 %	68 %	48 %	48 %	80 %	84 %
MIT [6]	50 %	-	-	-	64 %	-
SSP [7]	-	-	-	-	54 %	32 %
SVB frieze [8]	77 %	**89 %**	61 %	73 %	82 %	80 %
LBP-TOP	75 %	83 %	**75 %**	**83 %**	**88 %**	**88 %**

Dynamic texture synthesis

Guo Y, Zhao G, Chen J, Pietikäinen M & Xu Z (2009) Dynamic texture synthesis using a spatial temporal descriptor. Proc. IEEE International Conference on Image Processing (ICIP), 2277-2280.

- Dynamic texture synthesis is to provide a continuous and infinitely varying stream of images by doing operations on dynamic textures.

Introduction

- **Basic approaches to synthesize dynamic textures:**

  - parametric approaches
- physics-based
- method and image-based method

  - nonparametric approaches: they copy images chosen from original sequences and depends less on texture properties than parametric approaches

- **Dynamic texture synthesis has extensive applications in:**

  - video games
  - movie stunt
  - virtual reality

Synthesis of dynamic textures using a new representation

A. Schödl, R. Szeliski, D. Salesin, and I. Essa, “Video textures,” in Proc. ACM SIGGRAPH, pp. 489-498, 2000.

- The basic idea is to create transitions from frame i to frame j anytime the successor of i is similar to j, that is, whenever $D_{i+1,\, j}$ is small.

- **The algorithm of the dynamic texture synthesis:**

1. **Frame representation;**
2. **Similarity measure;**
3. Distance mapping;
4. Preserving dynamics;
5. Avoid dead ends;
6. Synthesis

Calculate the concatenated local binary pattern histograms from three orthogonal planes for each frame of the input video

Compute the similarity measure Dij between frame pair *Ii* and *I j* by applying Chi-square to the histogram of representation

To create transitions from frame *i* to *j* when *i* is similar to *j* , all these distances are mapped to probabilities through an exponential function Pij. The next frame to display after *i* is selected according to the distribution of *Pij*.

Match subsequences by filtering the difference matrix Dij with a diagonal kernel with weights [*w*−*m*,...,*wm*−1]

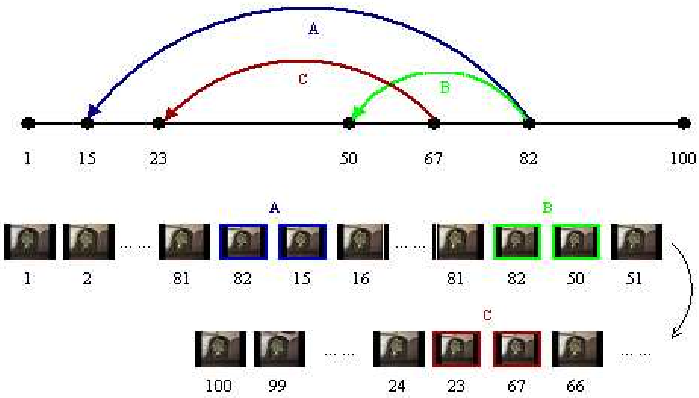
Distance measure can be updated by summing future anticipated costs

When transitions of video texture are identified, video frames are played by video loops

Synthesis of dynamic textures using a new representation

An example:

Considering that there are three transitions: $i_n \rightarrow j_n$ ($n = 1, 2, 3$) , loops from the source frame *i* to the destination frame *j* would create new image paths, named as loops. A created cycle is shown as:



Transitions	(i_n, j_n)
n=1	(82, 15)
n=2	(82, 50)
n=3	(67, 23)

Experiments

- We have tested a set of dynamic textures, including natural scenes and human motions.
  (http://www.texturesynthesis.com/links.htm and DynTex database, which provides dynamic texture samples for learning and synthesizing.)

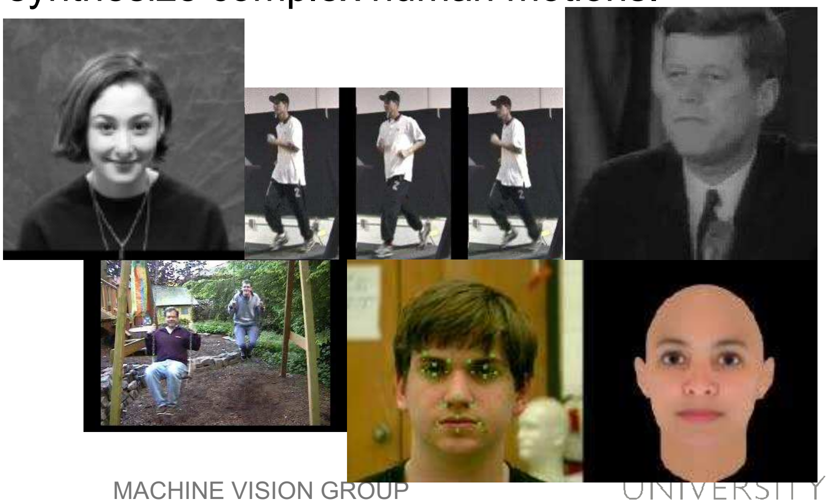
- The experimental results demonstrate our method is able to describe the DT frames from not only space but also time domain, thus can reduce discontinuities in synthesis.

Demo 1 i   Demo 1o   Demo 2 i   Demo 2 o

Experiments

- Dynamic texture synthesis of natural scenes concerns temporal changes in pixel intensities, while **human motion synthesis** concerns temporal changes of body parts.

- The synthesized sequence by our method maintains **smooth dynamic behaviors**. The better performance demonstrates its ability to synthesize complex human motions.



Examples of using LBP in different applications

Detection and tracking of objects

- Object detection [Zhang et al., IVC 2006]
- Human detection [Mu et al., CVPR 2008; Wang et al., ICCV 2009]
- On-line boosting [Grabner & Bishof, CVPR 2006]

Biometrics

- Fingerprint matching [Nanni & Lumini, PR 2008]
- Finger vein recognition [Lee et al., IJIST 2009]
- Touch-less palmprint recognition [Ong et al., IVC 2008]
- Gait recognition [Kellokumpu et al., 2009]
- Eye localization [Kroon et al., CVIU 2009]
- Face recognition in the wild [Wolf et al., ECCV 2008]
- Face verification in web image and video search [Wang et al., CVPR 2009]

Visual inspection

- Paper characterization [Turtinen et al.., IJAMT 2003]
- Separating black walnut meat from shell [Jin et al., JFE 2008 2009]
- Fabric defect detection [Tajeripur et al., EURASIP JASP 2008]

Biomedical applications

- Cell phenotype classification [Nanni & Lumini, Artif. Intell. Med. 2008]
- Diagnosis of renal cell carcinoma [Fuchs et al., MICCAI 2008]
- Ulcer detection in capsule endoscope images [Li & Meng, IVC 2009]
- Mass false positive reduct. in mammographic images [Llado et al., CMIG 2009]
- Lung texture analysis in CT images [Sorensen et al., IEEE TMI 2010]
- Retrieval of brain MR images [Unay et al., IEEE TITM 2010]
- Quantitative analysis of facial paralysis [He et al., IEEE TBE 2009]

Video analysis, photo management, and interactive TV

- Concept detection [Wu et al., ICIP 2008; Li et al., CVPR 2008]
- Overlay text detection and extraction from videos   [Kim & KIM, IEEE TIP 2009]
- Crowd estimation [Ma et al., ISIITA 2008]
- EasyAlbum – interactive photo annotation system [Cui et al.,CM CHI 2007]
- Cognitive face analysis for interactive TV [Ho An & Chung, IEEE TCE 2009]

Part 5: Summary and some future directions

Matti Pietikäinen

Summary

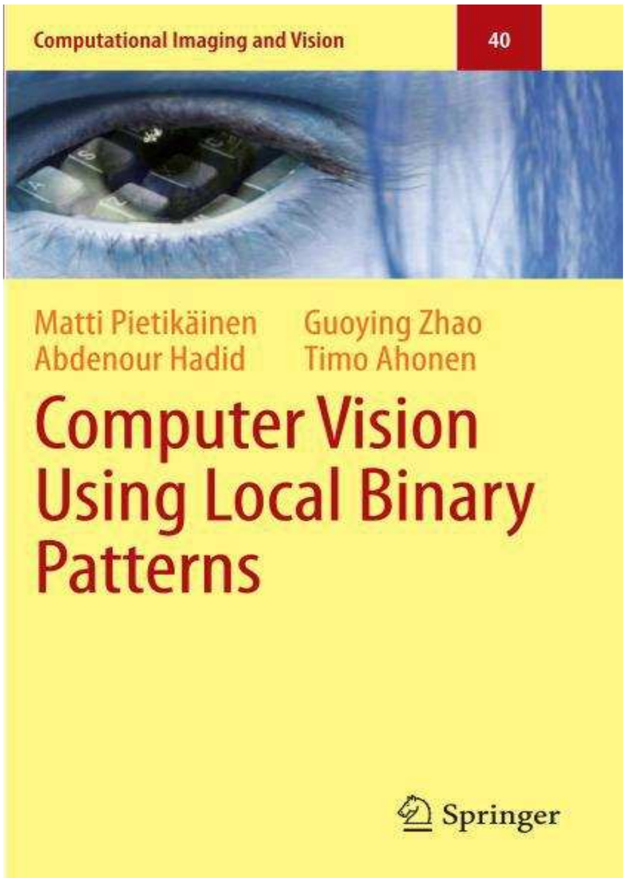
- Modern texture operators form a generic tool for computer vision
- LBP and its variants are very effective for various tasks in computer vision
- The choice between LBP and LPQ depends on the problem at hand

- The advantages of the LBP and its variants include
  - computationally very simple
  - can be easily tailored to different types of problems
  - robust to illumination variations
  - robust to localization errors
- LPQ is also insensitive to image blurring
- For a bibliography of LBP-related research, see
  http://www.cse.oulu.fi/MVG/LBP_Bibliography

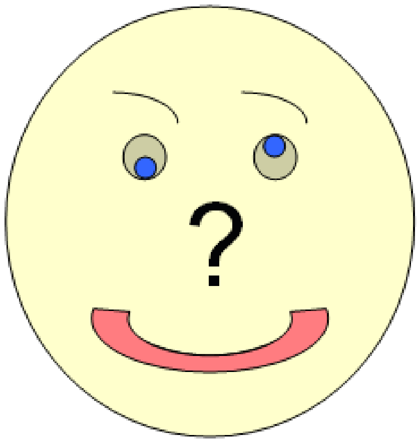
Some future directions

- New LBP/LPQ variants are still emerging
- Often a single descriptor is not effective enough
- Multi-scale processing
- Use of complementary descriptors
  - CLBP, Gabor&LBP, SIFT&LBP, HOG&LBP, LPQ&LBP
- Combining local with more global information (e.g. LBP & Gabor)
- Combining texture and color
- Combining sparse and dense descriptors
- Machine learning for finding the most effective descriptors for a given problem
- Dynamic textures offer a new approach to motion analysis
  - general constraints of motion analysis (i.e. scene is Lambertian, rigid and static) can be relaxed

A book on LBP will be published by fall 2011



Infotech Oulu
MACHINE VISION GROUP
UNIVERSITY of OULU
OULUN YLIOPISTO



Thanks!