TUTORIAL

ICCV 2009
September 27, 2009

Local Texture Descriptors in Computer Vision

Prof. Matti Pietikäinen, Dr. Guoying Zhao
{mkp,gyzhao}@ee.oulu.fi

Machine Vision Group
University of Oulu, Finland
http://www.ee.oulu.fi/mvg/

Texture is everywhere: from skin to scene images

Contents

Part 1: Milestones in texture research

ICCV 2009
September 27, 2009

Matti Pietikäinen
mkp@ee.oulu.fi

Taxonomies of texture

- Microtextures vs. macrotextures
- Stochastic (or irregular or random) vs. deterministic (or regular or structured)
- Coarseness, directionality, contrast, line-likeness, regularity and roughness (Tamura et al. 1978)
- Uniformity, density, coarseness, roughness, regularity, linearity, directionality, direction, frequency, and phase (Laws 1980)
- Three orthogonal dimensions of texture (Rao & Lohse 1993)
  - repetitive vs. non-repetitive
  - high contrast and non-directional vs. low-contrast and directional
  - granular, coarse and low-complexity vs. non-granular, fine and high complexity

Requirements for texture operators

Due to the variety of textures, we cannot expect that a single operator for texture description is adequate
- efficient discrimination of different types of textures
- robustness to pose and scale variations
- robustness to illumination variations
- robustness to spatial nonuniformity
- should work well for fairly small sample sizes
- low computational complexity

Tuceryan and Jain (1993, 1999) divided texture operators into
- statistical,
- geometrical,
- model based, and
- signal processing methods

Graylevel difference statistics (A Rosenfeld & E Troy: 1970; J Weszka et al.: IEEE T-SMC, 1976)

- First order statistics of local property values (i.e., means, variances) was one of the first approaches for texture description
- Local properties based on absolute differences between pairs of gray levels or of average gray levels were quite powerful (gray level difference histograms)
- Gray level difference statistics are a subset of the co-coccurrence matrix
- Sum and difference histograms were used as features by M Unser, IEEE TPAMI, 1986

Graylevel co-occurrence matrices (R Haralick et al.: IEEE T-SMC, 1973)

Joint gray level distribution for two gray levels located at a specified distance and angle (second order statistics)

- Haralick derived a set of 14 moments

+ has been most widely used texture method
+ works very well for stochastic textures
- computationally expensive
- needs gray-scale normalization / requantization

0	0	1	1
0	0	1	1
0	2	2	2
2	2	3	3

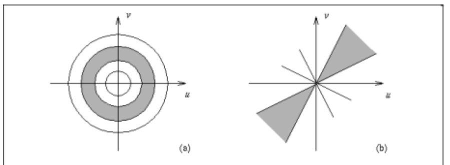
Gray level image.

$$P_{0^\circ,1} = \begin{vmatrix} 4 & 2 & 1 & 0 \\ 2 & 4 & 0 & 0 \\ 1 & 0 & 6 & 1 \\ 0 & 0 & 1 & 2 \end{vmatrix} \qquad P_{135^\circ,1} = \begin{vmatrix} 2 & 1 & 3 & 0 \\ 1 & 2 & 1 & 0 \\ 3 & 1 & 0 & 2 \\ 0 & 0 & 2 & 0 \end{vmatrix}$$

Generalized cooccurrence matrices were proposed by LS Davis et al., IEEE TPAMI,1981

MACHINE VISION GROUP
UNIVERSITY of OULU
OULUN YLIOPISTO

Fourier analysis (R Bajcsy: CGIP, 1973)



Partitioning of Fourier spectrum: (a) Ring filter, (b) wedge filter reflecting the Fourier spectrum symmetry.

MACHINE VISION GROUP
UNIVERSITY of OULU
OULUN YLIOPISTO

Laws' texture energy measures (K Laws: PhD thesis, 1980)

Masks for computing average gray level, edges, spots, ripples, waves

L3 = (1,2,1) center-weighted averaging
E3 = (-1,0,1) first difference – edge detection
S3 = (-1,2,-1) second difference – spot detection

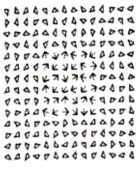
$$\begin{aligned} L_5 &= (1,4,6,4,1) \\ E_5 &= (-1,-2,0,2,1) \\ S_5 &= (-1,0,2,0,-1) \\ R_5 &= (1,-4,6,-4,1) \\ W_5 &= (-1,2,0,-2,-1) \end{aligned} \qquad L_5^T \times S_5 = \begin{bmatrix} -1 & 0 & 2 & 0 & -1 \\ -4 & 0 & 8 & 0 & -4 \\ -6 & 0 & 12 & 0 & -6 \\ -4 & 0 & 8 & 0 & -4 \\ -1 & 0 & 2 & 0 & -1 \end{bmatrix}$$

MACHINE VISION GROUP
UNIVERSITY of OULU
OULUN YLIOPISTO

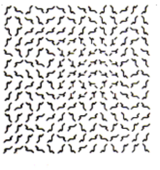
Julesz's textons (B Julesz: Nature, 1981)

Color
Terminator, number of end-of-lines. Ex. Closure, Connectivity
Elongated blobs of different sizes. Ex. Granularity

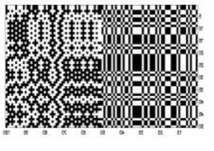
Closure

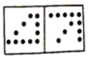


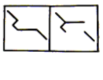
Connectivity



Granularity



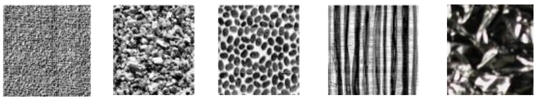




MACHINE VISION GROUP
UNIVERSITY of OULU
OULUN YLIOPISTO

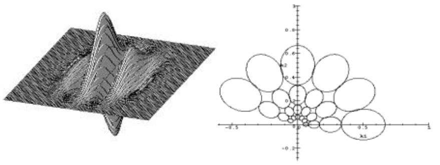
Textural features based on human perception (H Tamura et al.: IEEE T-SMC, 1978)

- Coarseness, contrast, directionality, line-likeness, regularity, roughness
- Features based on these ideas are used in some content-based image retrieval systems



MACHINE VISION GROUP
UNIVERSITY of OULU
OULUN YLIOPISTO

Gabor filters (M Turner: Biol. Cybern., 1986; M Clark & A Bovik: PRL, 1987; AK Jain & F Farrokhnia, PR, 1991; BS Manjunath & WY Ma: IEEE TPAMI, 1996)



+ biologically inspired
+ minimizes the joint uncertainty in space and frequency
+ supports both frequency analysis and spatial pattern approaches (large masks -> frequency analysis; small masks -> texture element detectors)
- selection of optimal features is not so straightforward
- computationally expensive
- does not work so well for stochastic textures

Wavelets : FS Cohen et al., IEEE TPAMI, 1993; T Chang & CCJ Kuo, IEEE T-IP, 1993; M Unser, IEEE T-IP, 1995; etc.

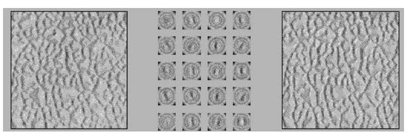
(Gaussian) random field models (G. Cross & A Jain: IEEE TPAMI, 1983; R Chellappa & S Chatterjee: ASSP, 1985; S Zhu et al.: IJCV, 1997)

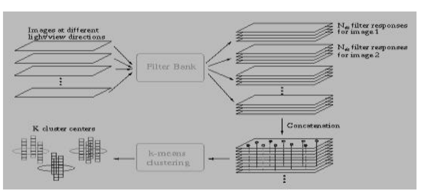
Markov Random Field is a conditional probability model allowing to model local spatial interactions among pixels

Basic principle: Detemine model to describe texture
Use model parameters for classification

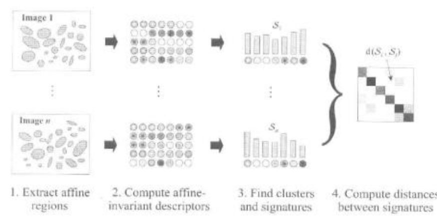
+ theoretically elegant
+ straightforward to use for texture synthesis
- sensitive to gray level distortions
- cannot model all textures
- computationally expensive

2D and 3D textons (J Malik et al.: ICCV 1999; T Leung & J Malik: IJCV, 2001; OG Cula & KJ Dana: CVPR 2001; M Varma & A Zisserman: ECCV, 2002; SC Zhu et al.: IJCV, 2005)





Sparse affine-invariant texture descriptors (S Lazebnik et al.: IEEE TPAMI, 2005)



+ modern approach based on interest region descriptors
+ robust to affine image transformations
- requires quite large windows
- does not work well for stochastic textures
- computationally expensive

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

Part 2: Local binary pattern (LBP) operators in spatial domain

ICCV 2009
September 27, 2009

Matti Pietikäinen
mkp@ee.oulu.fi

Contents

2.1 Theoretical foundations of LBP operators

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**

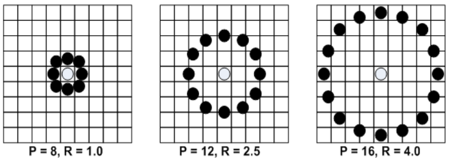
C = (6+7+8+9+7)/5 - (5+2+1)/3 = **4.7**

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

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.

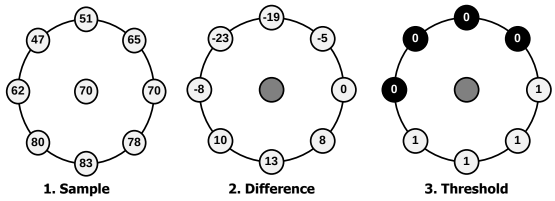
- arbitrary circular neighborhoods
- uniform patterns
- multiple scales
- rotation invariance
- gray scale variance as contrast measure



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:

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



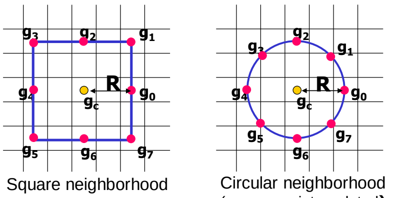
1. Sample **2. Difference** **3. Threshold**

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

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):

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

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:

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

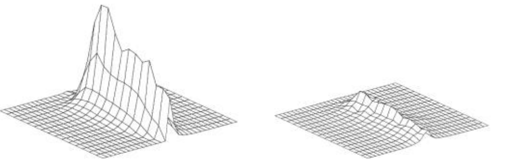
$t(g_c)$ describes the overall luminance of the image, which is unrelated to local image texture, hence we ignore it:

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

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]

Signed gray level differences

Ojala T, Valkealahti K, Oja E & Pietikäinen M (2001) Texture discrimination with multidimensional distributions of signed gray level differences. Pattern Recognition 34:727-739.

Cooccurring differences provide more information than just one. Computing cooccurring differences in 3x3 subimages:

g_3	g_2	g_1
g_4	g_c	g_0
g_5	g_6	g_7

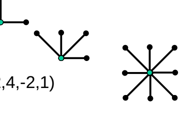
we estimate distributions

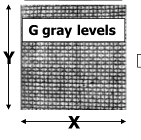
$p_2\,(g_0 - g_c, g_2 - g_c)$
$p_4\,(g_0 - g_c, g_1 - g_c, g_2 - g_c, g_3 - g_c)$
$p_8\,(g_0 - g_c, g_1 - g_c, \ldots, g_7 - g_c)$

Signed gray level differences (cont.)

1	2	5
7	5	8
9	3	6

$p_2 = (3,-3)$
$p_4 = (3,-3,0,-4)$
$p_8 = (3,-3,0,-4,2,4,-2,1)$





(Y-2)(X-2) difference vectors

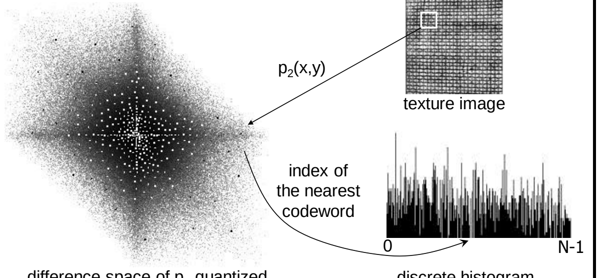
p_k (k=2,4,8) estimated with a discrete histogram of N bins

Stability criterion: $f_{ave} = (Y-2)(X-2)/N \geq f_{min}$ (~ 5 or 10)

Volume of the difference space: $V = (2G-1)^k >> N$

Signed gray level differences (cont.)

Vector quantization of the difference space



difference space of p_2 quantized with a codebook of N codewords $N \sim (X-2)(Y-2)/f_{min}$

discrete histogram estimating p_2

LBP: Local Binary Pattern

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

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

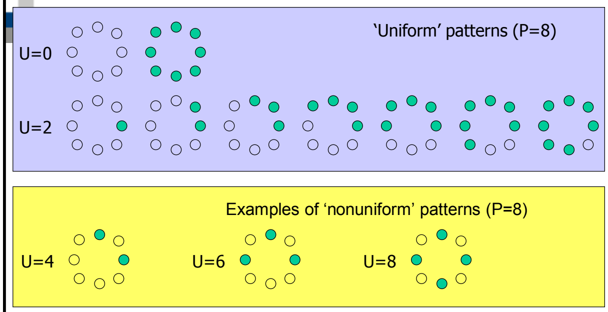
where

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

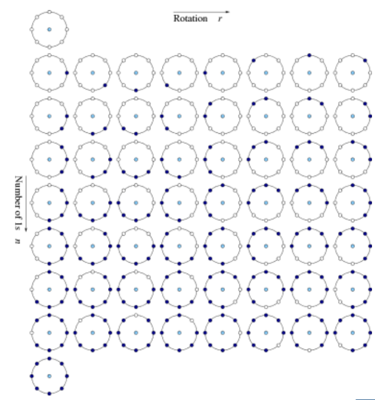
Above is transformed into a unique P-bit pattern code by assigning binomial coefficient 2^p to each sign $s(g_p - g_c)$:

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

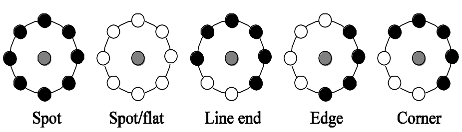
‘Uniform’ patterns



- 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

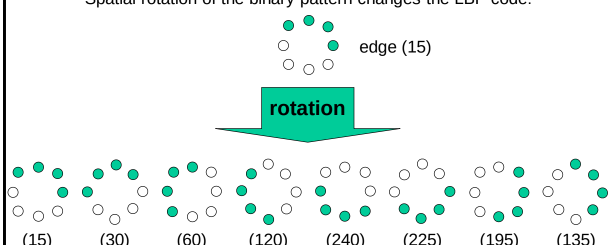


Texture primitives (“micro-textons”) detected by the uniform patterns of LBP



Rotation of Local Binary Patterns

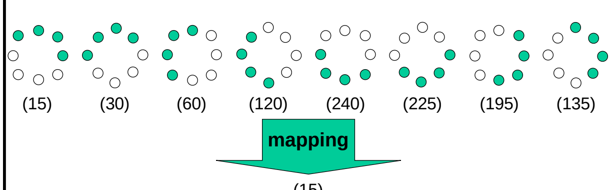
Spatial rotation of the binary pattern changes the LBP code:



Rotation invariant local binary patterns

Formally, rotation invariance can be achieved by defining:

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



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

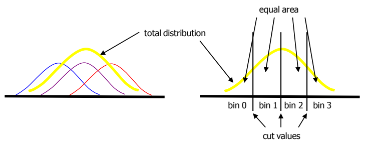
$$m = \frac{1}{P} \sum_{p=0}^{P-1} g_p$$

$VAR_{P,R}$
- invariant wrt. gray scale shifts
- invariant wrt. rotation along the circular neighborhood

Quantization of continuous feature space

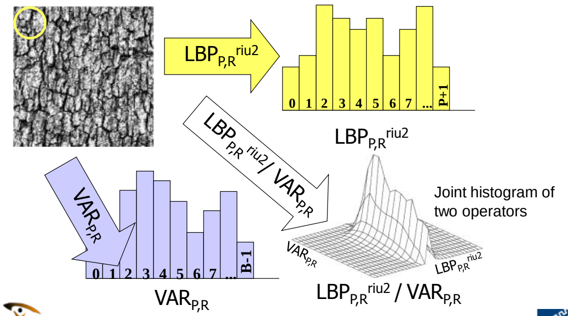
Texture statistics are described with discrete histograms
- Mapping needed for continuous-valued contrast features

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



Multiscale analysis

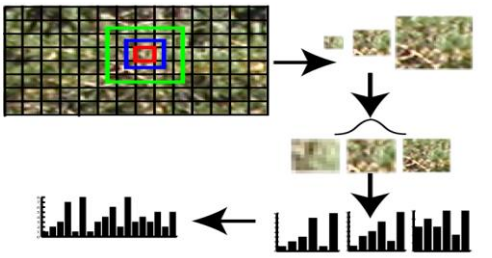
Information provided by N operators can be combined simply by summing up operatorwise similarity scores into an aggregate similarity score:

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

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

Multi-scale analysis (cont.)

Image regions can also be re-scaled prior to feature extraction



Nonparametric classification principle

Sample S is assigned to the class of model M that maximizes

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

Many other dissimilarity measures can be used (chi square, histogram intersection, Kullback-Leibler divergence, Jeffrey's divergence, etc.)

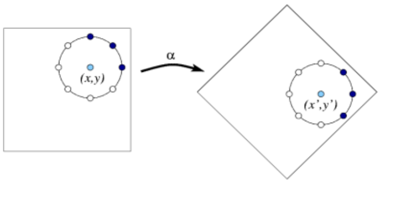
Nonparametric: no assumptions about underlying feature distributions are made!!

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.

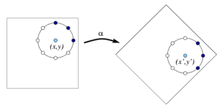
Rotation revisited

Rotation of an image by α degrees
- Translates each local neighborhood to a new location
- Rotates each neighborhood by α degrees



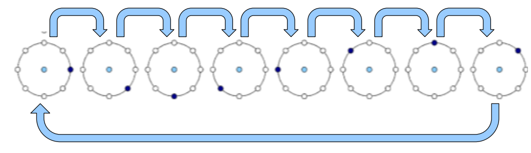
Rotation revisited (2)

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 (3)

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



Rotation invariant features

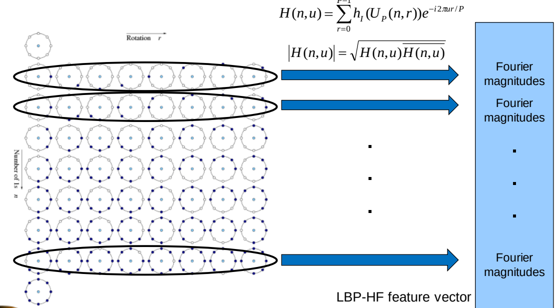
LBP histogram features that are *invariant to cyclic shifts* along the rows are invariant to *k**45˚ rotations of the input image

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

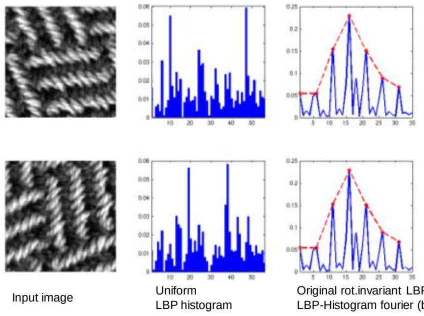
LBP Histogram Fourier Features

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

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

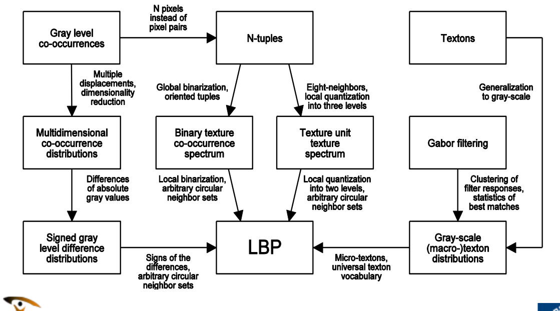


Example



Input image
Uniform LBP histogram
Original rot.invariant LBP (red)
LBP-Histogram fourier (blue)

LBP unifies statistical and structural approaches



2.2 LBP in various computer vision problems

LBP has become very widely used in various computer vision problems and applications due to its high discriminative power, tolerance against illumination changes and computational simplicity.

Next we give examples of our research applying spatial domain LBP to unsupervised segmentation, recognition of 3D textured surfaces, image retrieval, face analysis, description of interest regions, and modeling the background and detecting moving objects.

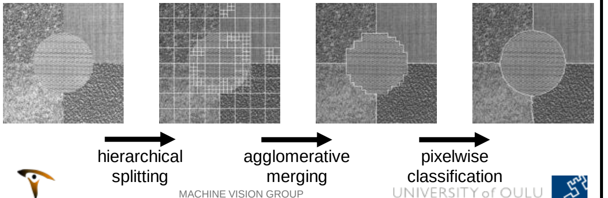
Unsupervised texture segmentation

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

- LBP/C was used as texture operator

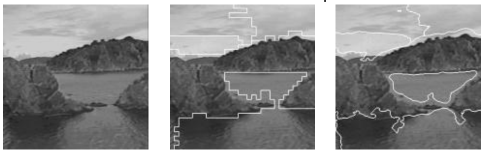
Segmentation algorithm consists of three phases:

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

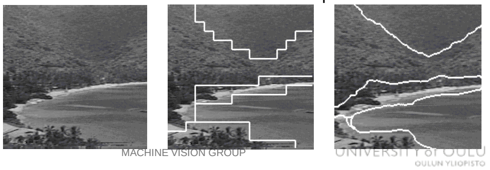


Examples of segmentation

Natural scene #1: 384x384 pixels



Natural scene #2: 192x192 pixels



View-based recognition of 3D-textured surfaces

Pietikäinen M, Nurmela T, Mäenpää T & Turtinen M (2003) View-based recognition of real-world textures. Pattern Recognition 7(32):313-323.

Due to the changes in viewpoint and illumination, the visual appearance of different surfaces can vary greatly

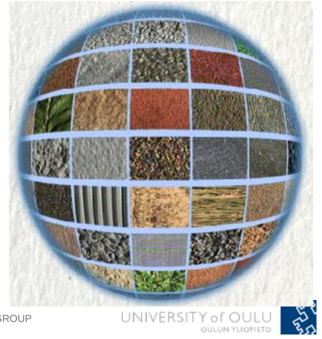
- textures are modeled with multiple histograms of micro-textons extracted with the LBP operator
- provided the leading performance in the classification of CUReT textures taken from different view angles and illuminations
- very promising results in the classification of outdoor scene images
- an approach to learning appearance models for view-based texture recognition using self-organization of feature distributions was also proposed

Classification results for CUReT textures

Each image class included 118 images taken under varying viewpoint and illumination

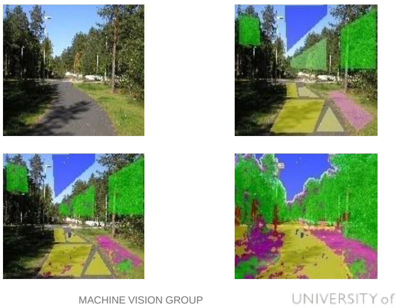
Classification rates:
- 98.8% for 20 classes
- 97.3% for 40 classes
- 96.6% for 61 classes

Results were better than those obtained, e.g., by Cula & Dana (2001) and Varma and Zisserman (2002)



Example of view-based classification of an Outex image

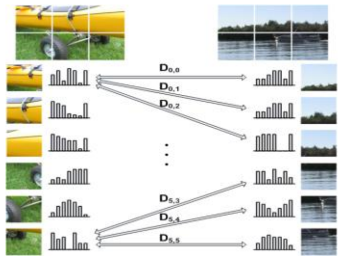
(a) The original image (b) Ground-truth regions
(c) Classified pixels within ground-truth regions (d) Segmented image



LBP in image retrieval

Takala V, Ahonen T & Pietikäinen M (2005) Block-based methods for image retrieval using local binary patterns. In: Image Analysis, SCIA 2005 Proceedings, Lecture Notes in Computer Science 3540, 882-891.

A block division method for content-based retrieval (best results are obtained with overlapping blocks)

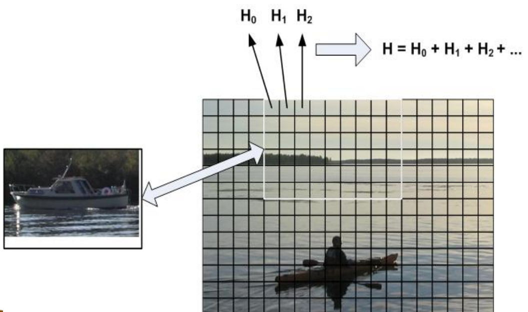


Experiments with images from Corel Gallery database

- 27 categories with 50 images in each were used
- Block based LBP method performed better than Edge Histogram (of MPEG-7) and color correlogram features

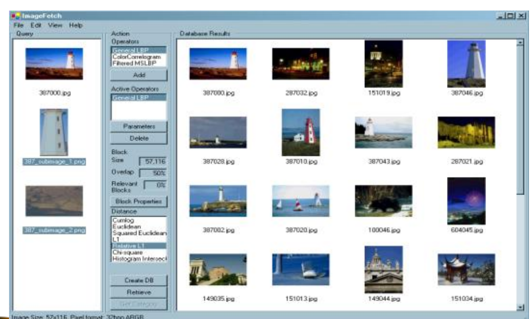
Block-based image retrieval

The primitive blocks approach



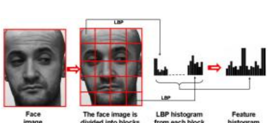
Block-based image retrieval

Query results with the primitive-based approach



Face analysis using local binary patterns

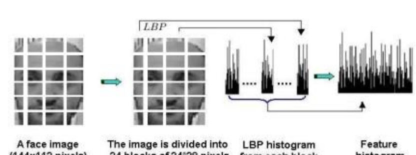
- Face recognition is one of the major challenges in computer vision
- We proposed (ECCV 2004, PAMI 2006) a face descriptor based on LBP's
- Our method has already been adopted by many leading scientists and groups
- Computationally very simple, excellent results in face recognition and authentication, 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:

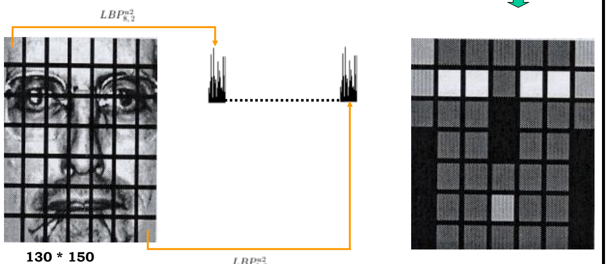


Weighting the regions

Block size 18 * 21

Metrics χ^2 (Chi-square)

Weighting



Feature vector length → 2891

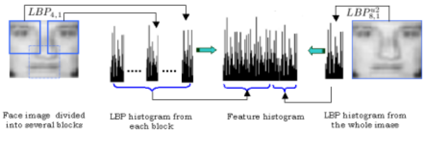
LBP in AuthenMetric F1

Institute of Automation, Chinese Academy of Sciences

Face detection with LBP

Hadid A, Pietikäinen M & Ahonen T (2004) A discriminative feature space for detecting and recognizing faces. Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2004), 2:797-804.

A facial description for small-sized face images:



Feature vector length → 203

Face detection results

Face detection results

Table 3.4. Comparative performance of detecting 227 faces in 80 images.

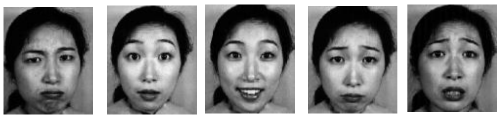
Method	Detected	False detection	Rates
Schneiderman-Kanade(1.0, 1.0)	218	41	96.0 %
BDF Method	221	1	97.4 %
Normalized Pixel features	213	6	93.8 %
$LBP_{4,1}$+$LBP^{u2}_{8,1}$ (203 bins)	**221**	**0**	97.4 %

Application example: FP7 project: Mobile Biometry (MOBIO) 2008-2010 (www.mobioproject.org)

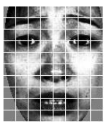
- The aim of is to investigate multiple aspects of biometric authentication based on the face and voice in the context of mobile devices
- To increase security and user acceptance - using standard sensors already available on mobile phones
- Coordinator: IDIAP Research Institute (CH)
- Partners: University of Manchester (UK), University of Surrey (UK), Universite d'Avignon (FR), Brno University of Technology (CZ), University of Oulu (FI), IdeArk (CH), EyePmedia (CH), Visidon (FI)

LBP in facial expression recognition from still images

Feng X, Pietikäinen M & Hadid A (2005) Facial expression recognítion with local binary partterns and linear programming. Pattern Recognition and Image Analysis 15(2):546-548.



Japanese Female Facial Expression database (JAFFE)

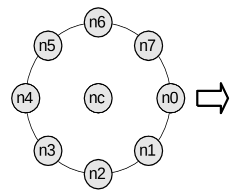


- Linear programming technique was adopted to classify seven facial expressions: anger, discust, fear, happiness, sadness, surprise, and neutral

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.

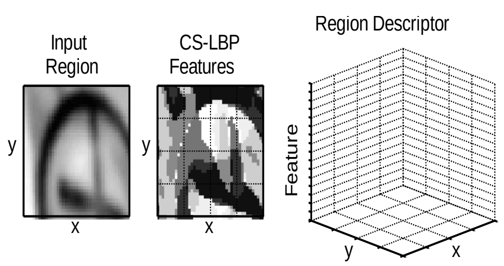
Neighborhood

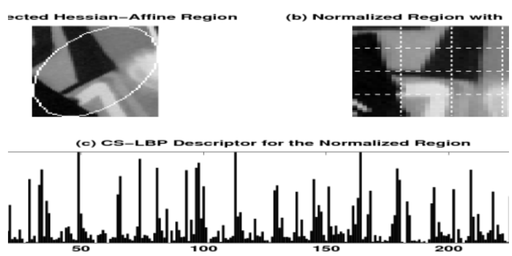


Binary Pattern

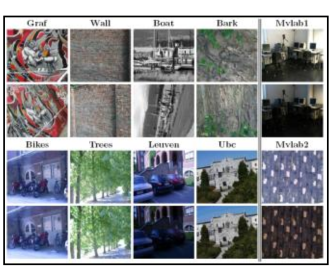
LBP =	CS-LBP =
$s(n0 - nc)2^0$ +	$s(n0 - n4)2^0$ +
$s(n1 - nc)2^1$ +	$s(n1 - n5)2^1$ +
$s(n2 - nc)2^2$ +	$s(n2 - n6)2^2$ +
$s(n3 - nc)2^3$ +	$s(n3 - n7)2^3$
$s(n4 - nc)2^4$ +	
$s(n5 - nc)2^5$ +	
$s(n6 - nc)2^6$ +	
$s(n7 - nc)2^7$	

Description of interest regions





Setup for image matching experiments



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

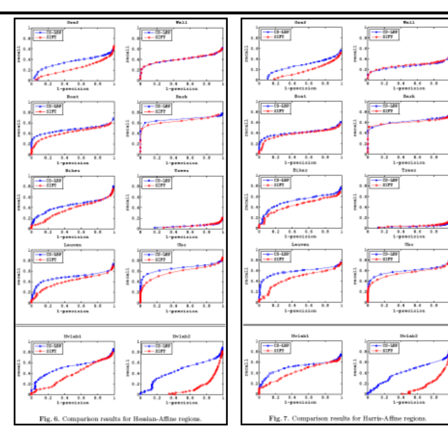
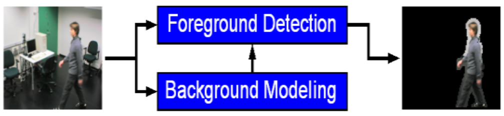


Fig. 6. Comparison results for Hessian-Affine regions.

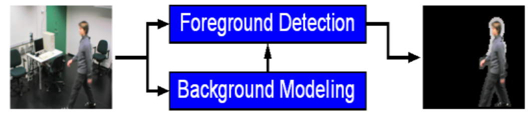
Fig. 7. Comparison results for Harris-Affine regions.

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.

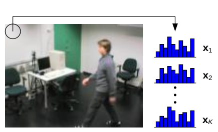
…Overview of the approach…

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.



…Overview of the approach… Background modeling

1. Calculate an LBP histogram $\mathbf{x}_t$ for the pixel of the new video frame.
2. Compare the new pixel histogram $\mathbf{x}_t$ against the existing *K* model histograms $\{\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_K\}$ by using the *histogram intersection* as the distance measure.
   1. If none of the model histograms is close enough to the new histogram, the model histogram with the lowest weight is replaced with the new histogram and is given a low initial weight.
   2. If a model histogram close enough to the new histogram was found, the bins of this histogram are updated as follows:
      $\mathbf{x}_{k,t}[i] = \alpha_b \mathbf{x}_t[i] + (1 - \alpha_b)\mathbf{x}_{k,t-1}[i] \qquad 0 \le \alpha_b \le 1 \qquad (1)$
      Furthermore, the weights of the model histograms, $\{\omega_1, \omega_2,\ldots, \omega_K\}$, are updated as follows:
      $\omega_{k,t} = (1 - \alpha_w)\, \omega_{k,t-1} + \alpha_w M_{k,t} \qquad 0 \le \alpha_w \le 1 \qquad (2)$ where $M_{k,t}$ is 1 for the matched histogram and 0 for the others.

…Overview of the approach… Background modeling

3. Decide which of the histograms of the model are most likely produced by the background processes. We use the persistence of the histogram as an evidence for this. Because the persistence of the k^{th} model histogram is directly related to its weight $\omega_{k,t}$, the model histograms are sorted in decreasing order according to their weights. As a result the most probable background histograms are on the top of the list.
4. As a last phase of the updating procedure, the first *B* model histograms are selected to be the background model as follows:
   $\omega_{1,t} + \omega_{2,t} + \ldots + \omega_{B,t} > T_B \qquad 0 \le T_B \le 1 \qquad (3)$

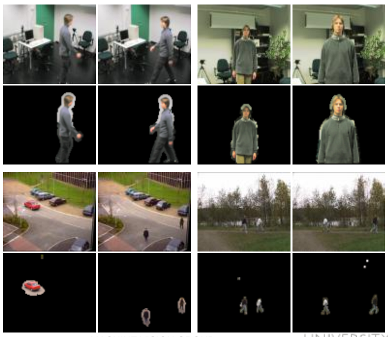
…Overview of the approach… Foreground detection

Foreground detection is achieved via comparison of the new pixel histogram $\mathbf{x}_t$ against the existing *B* background histograms $\{\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_B\}$ selected at the previous time instant.

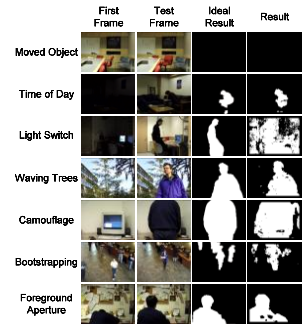
• If a match is **not** found, the pixel is considered to belong to the foreground.

• Otherwise, the pixel is marked as background.

Examples of detection results



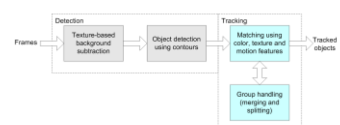
Detection results for images of Toyama et al. (ICCV 1999)



Demo for detection of moving objects

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.



Part 3: Motion analysis with spatiotemporal LBPs

ICCV 2009
Sep. 27, 2009

Guoying Zhao
gyzhao@ee.oulu.fi

Contents

3.1 Dynamic texture recognition

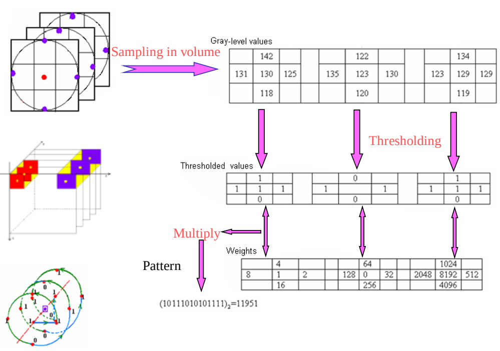
Zhao G & Pietikäinen M (2007) Dynamic texture recognition using local binary patterns with an application to facial expressions. IEEE Transactions on Pattern Analysis and Machine Intelligence 29(6):915-928. (parts of this were earlier presented at ECCV 2006 Workshop on Dynamical Vision and ICPR 2006)

Dynamic texture

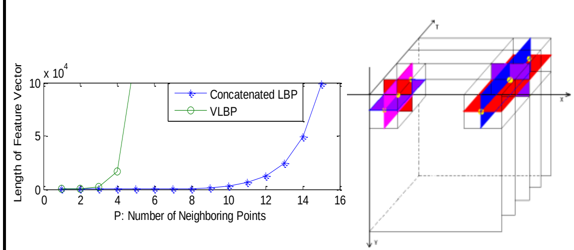
- Dynamic Textures (DT): Temporal texture
- Textures with motion
- An extension of texture to the temporal domain
- Encompass the class of video sequences that exhibit some stationary properties in time

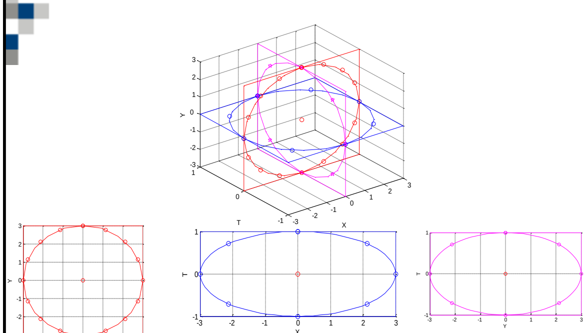
- Lots of dynamic textures in real world
- Description and recognition of DT is needed

Volume **L**ocal **B**inary **P**atterns (**VLBP**)

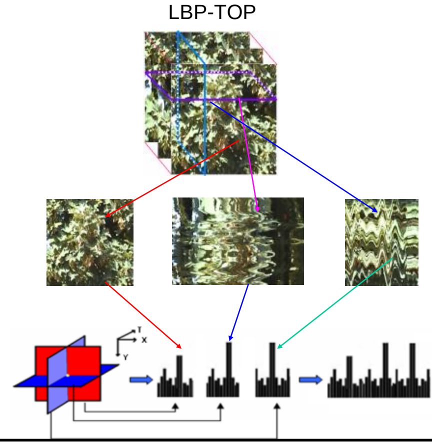


LBP from **T**hree **O**rthogonal **P**lanes (**LBP-TOP**)

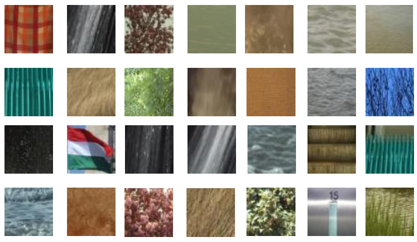




LBP-TOP

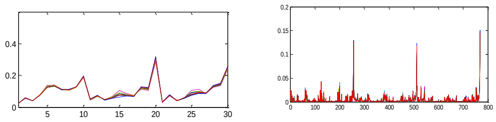


DynTex database



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

Results of LBP from three planes

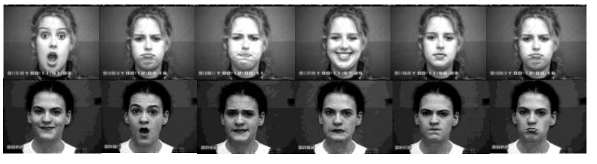


LBP	XY	XZ	YZ	Con	weighted
8,8,8,1,1,1 riu2	88.57	84.57	86.29	93.14	93.43[2,1,1]
8,8,8,1,1,1 u2	92.86	88.86	89.43	94.57	96.29[4,1,1]
8,8,8,1,1,1 Basic	95.14	90.86	90	95.43	97.14[5,1,2]
8,8,8,3,3,3 Basic	90	91.17	94.86	95.71	96.57[1,1,4]
8,8,8,3,3,1 Basic	89.71	91.14	92.57	94.57	95.71[2,1,8]

3.2 Facial expression 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.

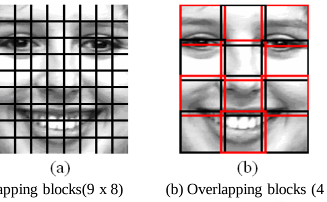
- Determine the emotional state of the face
  - Regardless of the identity of the face



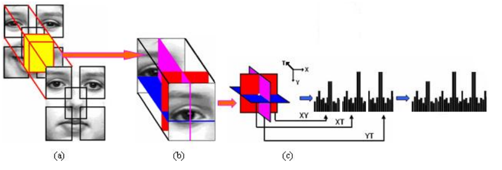
Facial Expression Recognition

- 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) Non-overlapping blocks(9 x 8) (b) Overlapping blocks (4 x 3, overlap size = 10)



(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

Angry

Disgust

Sadness

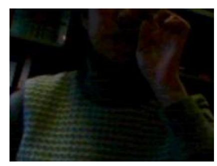
Fear

Surprise

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**

Example images in different illuminations

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

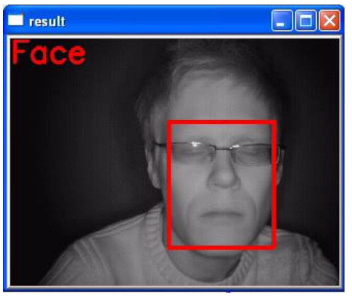
Strong illumination **Weak illumination** **Dark illumination**

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



Facial expression under NIR environment

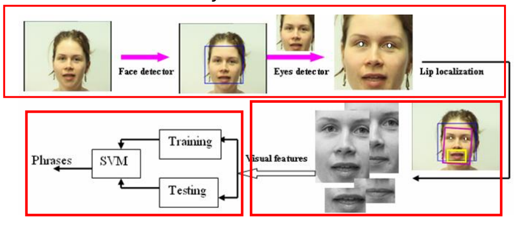
3.3 Visual speech recognition

Zhao G, Barnard M & Pietikäinen M (2009). Lipreading with local spatiotemporal descriptors. IEEE Transactions on Multimedia, in press.

- ❖ 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.

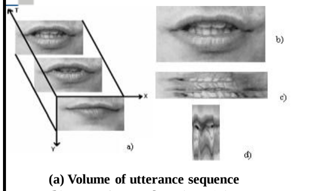
System overview



Our system consists of three stages.

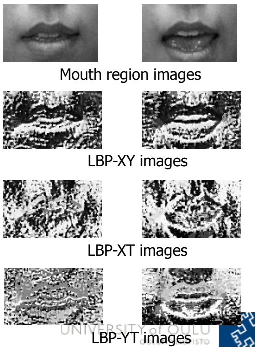
- First stage: face and eye detectors, and the localization of mouth.
- Second stage: extracts the visual features.
- Last stage: recognize the input utterance.

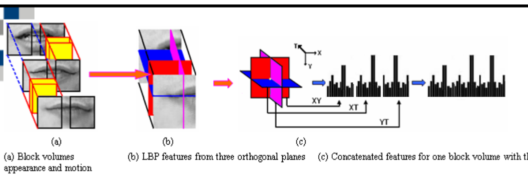
Local spatiotemporal descriptors for visual information



(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

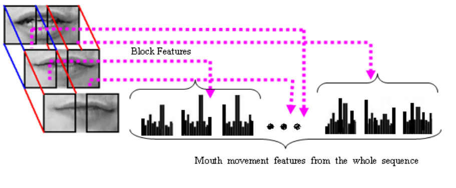
Overlapping blocks (1 x 3, overlap size = 10).





(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 representation.

Experiments

- Three databases:

1) 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"

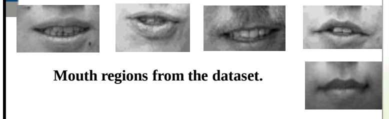
2) Tulips1 audio-visual database

12 subjects, pronouncing the first four digits in English two times in repetition. Totally 96 sequences.

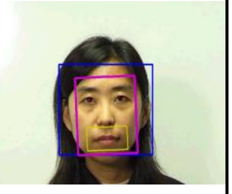
3) AVLetters database

10 people, each uttering 26 english letters three times. Totally 780 sequences.

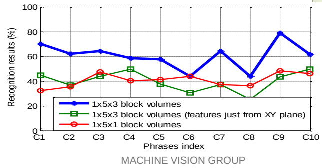
Experimental results - OuluVS database



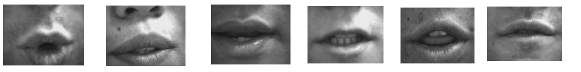
Mouth regions from the dataset.



Speaker-independent:



Experimental results - Tulips1 audio-visual database



Mouth images with translation, scaling and rotation from Tulips1 database.

Comparison to other methods on Tulips1 audio-visual database (speaker independent).

	Features	Normalization	Results (%)
[Arsic 2006]	MRPCA	Y	81.25
[Arsic 2006]	MI MRPCA	Y	87.5
[Gurban 2005]	Temporal Derivatives Features	Y	80 91(a&v, 10 dB SNR level)
Ours	$LBP-TOP_{8,8,8,1,1,1}$ Blocks: 3x6x2	**N**	**92.71**

Visemes	Phonemes	Visemes	Phonemes
/p/	P, B, M	/iy/	IY
/f/	F	/aa/	AA
/t/	T, D, S, Z	/ah/	AY
/ch/	CH, JH, ZH	/ow/	OW
/w/	W, R	/uw/	UW
/k/	K, N, L	/ey/	EH, EY

AVLetters database: 26 letters, 10 people, three utterances per letter.

Confusion matrix from SVMs($LBP-TOP^{u2}_{8,8,8,3,3,3}$ features with $2\times5\times3$ blocks)

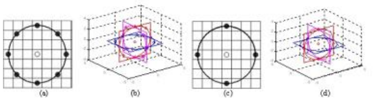
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	17	1			1				1		1	3		5					1							
B	1	18														10						1				
C			11	8	1		1				1					1				5		1				1
D			6	12			1				1									9						1
E	3			1	19						2								2	3						
F						25					1	1	1						1					1		
G			1				16			6										3		2			1	1
H						1		20	1		1	2		1					2		1			1		
I	3					1			18		1	1		2	1			3								
J			1				3			22	1									1		2				
K	2		1	2	1					1	17	1		2						3						
L	4				1			1			1	13	1	5				1		1				2		
M	1					1					1	1	23	1						1						1
N	5			1	2				2			3	1	11		1		1	2					1		
O		1													24		2	1			2					
P	2	10				1								1		15							1			
Q								1							1		17	1			10					
R	1						1		1		1				2			23							1	
S								1						2					19					8		
T			4	4	2						1									19						
U				1					1						3		12				13					
V			3							1	1											23				2
W																					1		28		1	
X						1		1			2			2					6	1				17		
Y		1															1								28	
Z			3	1						2									1	1						22

Principal appearance and motion from boosted spatiotemporal descriptors

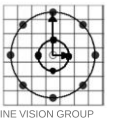
Zhao G & Pietikäinen M (2009) Boosted multi-resolution spatiotemporal descriptors for facial expression recognition. Pattern Recognition Letters 30(12):1117-1127.

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



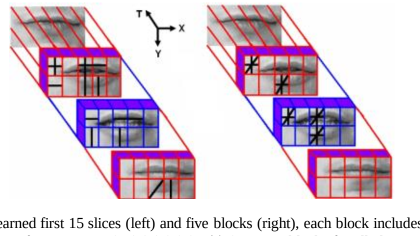
- 2) Use of different radii which can catch the occurrences in different space and time scales



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

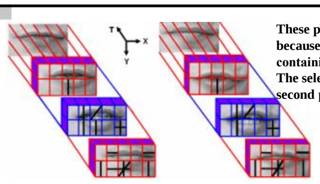
The first two resolutions focus on the
❖ pixel level in feature computation, providing different local spatiotemporal information

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



Learned first 15 slices (left) and five blocks (right), each block includes three slices from LBP − TOP8,8,8,3,3,3 with 2 × 5 × 3 blocks for all classes learning.

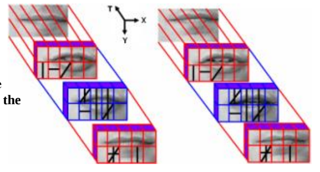
The selected features for all classes are mainly from YT slices (seven out of 15) and XT slices (seven out of 15), just one from XY slices. That suggests that in visual speech recognition the motion information is more important than the appearance.



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,

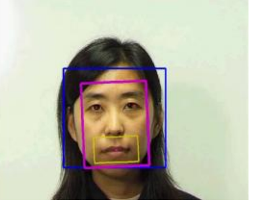
Selected 15 slices for phrases "See you" and "Thank you".

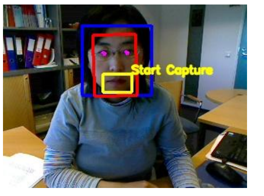
The phrases "excuse me" and "I am sorry"are different throughout the whole utterance, and the selected features also come from the whole pronunciation.



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

Demo for visual speech recognition





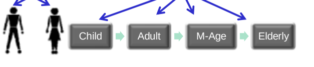
3.4 Face analysis from videos

Hadid A, Pietikäinen M & Li SZ (2007) Learning personal specific facial dynamics for face recognition from videos. Proc. 2007 IEEE International Workshop on Analysis and Modeling of Faces and Gestures (AMFG), 1-15.

Problem description

ID

How to efficiently recognize faces, determine gender, estimate age etc. from video sequences?



Traditional approaches..

The most common approach is to apply still image based methods to some selected (or all) frames

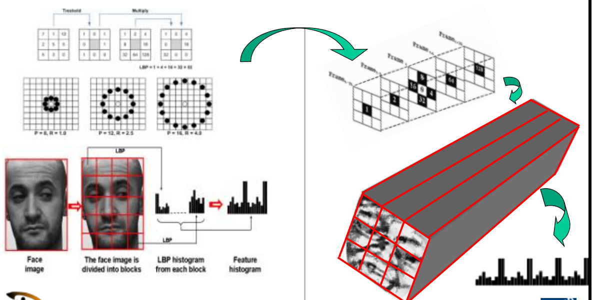
One new direction..

- A Spatiotemporal Approach to Face Analysis from Videos

Motivations:

neuropsychological studies indicating that facial dynamics do support face and gender recognition especially in degraded viewing conditions such as poor illumination, low image resolution…

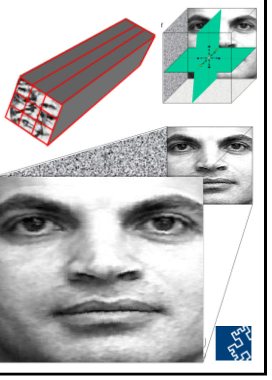
A face sequence can be seen as a collection of rectangular prisms (volumes) from which we extract local histograms of *Extended* Volume Local Binary Pattern code occurrences.



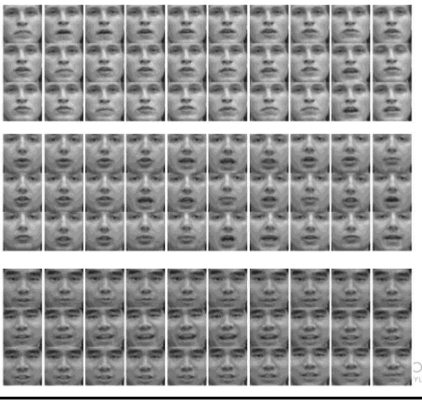
A spatiotemporal approach to face analysis from videos..

Algorithm:

1. **Divide the video into local prisms**
2. **Consider 3D neighborhood of each pixel**
3. **Apply VLBP**
4. **Feature Selection using AdaBoost**
5. **Extract local histograms**
6. **Histogram concatenation & normalization**
7. **Matching**



Some experimental results



Experiments on face recognition

Hadid A, Pietikäinen M & Li SZ (2007) Learning personal specific facial dynamics for face recognition from videos. Proc. 2007 IEEE International Workshop on Analysis and Modeling of Faces and Gestures (AMFG), 1-15.

Method	Results on MoBo	Results on Honda/UCSD	Results on CRIM
PCA	87.1%	69.9%	89.7%
LDA	90.8%	74.5%	91.5%
LBP [13]	91.3%	79.6%	93.0%
HMM [8]	92.3%	84.2%	85.4%
ARMA [7]	93.4%	84.9%	80.0%
VLBP [14]	90.3%	78.3%	88.7%
VLBP+AdaBoost	96.5%	89.1%	94.4%
EVLBP+AdaBoost	**97.9%**	**96.0%**	**98.5%**

Static image based versus spatiotemporal based approaches to face recognition

Experiments on gender classification

Hadid A & Pietikäinen M (2009) Combining appearance and motion for face and gender recognition from videos. Pattern Recognition 42:2818-2827.

Databases: CRIM, VidTIMIT and Cohn-Kanade

Gender classification results on test videos of familiar (columns 1-3) and unfamiliar subjects (columns 4-6). The methods are based on appearance only (1st, 2nd & 3rd rows), motion only (4th & 5th rows), and combination of appearance and motion (6th & 7th rows).

Method	Gender Classification Rate					
	Subjects Seen during Training			Subjects Unseen during Training		
	20×20	40 × 40	60×60	20×20	40 × 40	60×60
Pixels+SVM+Voting	93.1	93.3	91.9	88.5	89.4	88.2
LBP+SVM+Voting	94.0	94.4	95.4	90.1	90.6	91.0
XY-LBP+SVM	96.1	97.2	97.1	**95.5**	**95.7**	**96.3**
YT-LBP+SVM	74.5	81.6	83.2	51.6	49.7	50.4
XT-LBP+SVM	78.5	79.4	80.4	45.9	47.1	44.2
VLBP+SVM	98.2	98.3	98.8	82.7	84.3	84.7
EVLBP+AdaBoost	**100**	**100**	**100**	79.2	81.5	78.6

3.5 Activity recognition

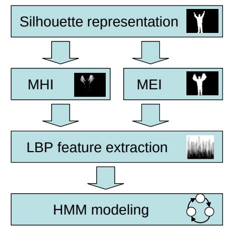
Kellokumpu V, Zhao G & Pietikäinen M (2009) Recognition of human actions using texture, a journal article in revision.

Texture based description of movements

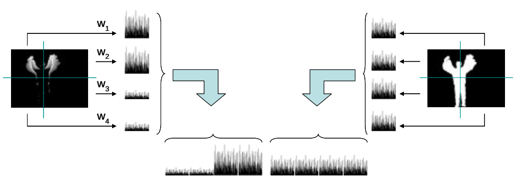
- We want to represent human movement with it's local properties
  - > Texture
- But texture in an image can be anything? (clothing, scene background)
  - > Need preprocessing for movement representation
  - > We use temporal templates to capture the dynamics
- We propose to extract texture features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

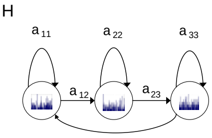


Features



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%

Experiments – Continuous data

- Detection and recognition experiments on database 1 using a sliding window based detection.
- Demo

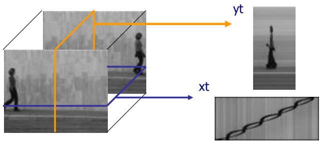
Activity recognition using dynamic textures

- 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.

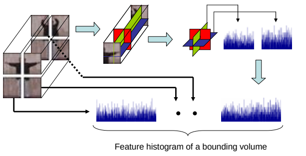
Dynamic textures for action recognition

- 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

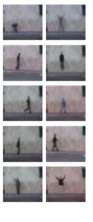


- HMM is used for sequential modeling

Action classification results – Weizmann dataset

- Classification accuracy 95,6% using image data

	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

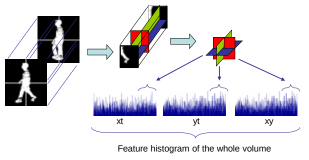


Action classification results - KTH

- 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



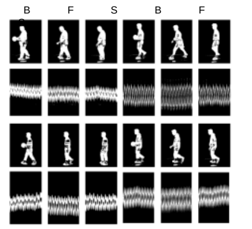
$$Similarity = \sum \min(h_i, h_j)$$

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~~)

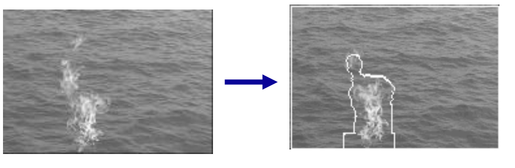


Experiments - Gait recognition results

	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 %**

3.6 Unsupervised dynamic texture segmentation

Chen J, Zhao G & Pietikäinen M (2008) Unsupervised dynamic texture segmentation using local spatiotemporal descriptors. Proc. International Conference on Pattern Recognition (ICPR), 4 p.



Input Output

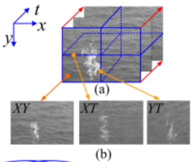
Dynamic texture segmentation

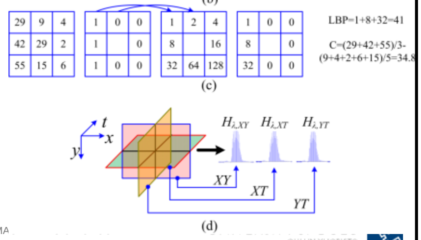
- Potential applications: Remote monitoring and various type of surveillance in challenging environments:
  - monitoring forest fires to prevent natural disasters
  - traffic monitoring
  - homeland security applications
  - animal behavior for scientific studies.

Related work

- Mixtures of dynamic texture model
  - A.B. Chan and N. Vasconcelos, PAMI2008
- Mixture of linear models
  - L. Cooper, J. Liu and K. Huang, Workshop in ICCV2005
- Multi-phase level sets
  - D. Cremers and S. Soatto, IJCV2004
- Gauss-Markov models and level sets
  - G. Doretto, A. Chiuso, Y. N. Wu and S. Soatto, ICCV2003
- Ising descriptors
  - A. Ghoreyshi and R. Vidal, ECCV2006
- Optical flow
  - R. Vidal and A. Ravichandran, CVPR2005

- Feature: $(LBP/C)_{TOP}$
  - Local binary patterns
  - Contrast
  - three orthogonal planes

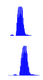




Measure

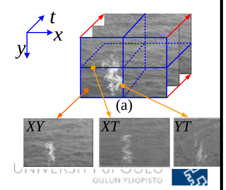
- Similarity measurement

$$\Pi(H_1, H_2) = \sum_{i=1}^{L} \min(H_{1,i}, H_{2,i})$$



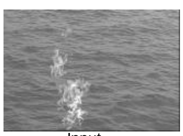
- Distance between two sub-blocks

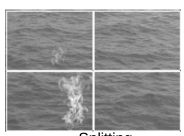
$\boldsymbol{d}=\{\Pi_{LBP,\,XY}, \Pi_{LBP,\,XT}, \Pi_{LBP,\,YT}, \Pi_{C,\,XY}, \Pi_{C,\,XT}, \Pi_{C,\,YT}\}^T.$

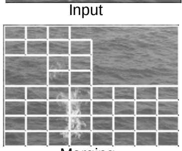


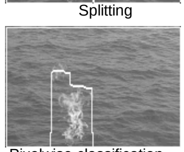
DT segmentation

- Three phases:
  Splitting, Merging, Pixelwise classification.

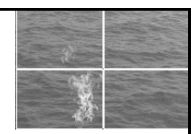

Input


Splitting

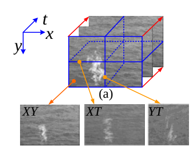

Merging


Pixelwise classification

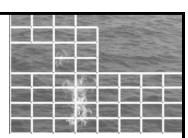
Splitting



- Recursively split each input frame into square blocks of varying size.
- criterion of splitting:
  - *one* of the features in the three planes (i.e., LBPπ and Cπ, π=*XY*, *XT*, *YT*) votes for splitting of current block

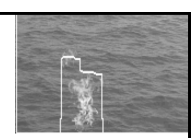


Merging



- Merge those similar adjacent regions with smallest merger importance (*MI*) value
- *MI* : *MI*=*f*(*p*)×(1-Π)
  - Π is the distance between two regions
  - *f*(*p*)= sigmoid(*βp*). (*β*=1, 2, 3, …)
    - $p=N_b/N_f$
    - N_b is the number of pixels in current block
    - N_f is the number of pixels in current frame

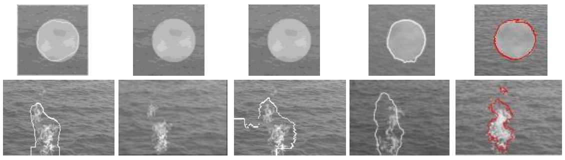
Pixelwise classification



- Compute $(LBP/C)_{TOP}$ histograms over its circular neighbor for each boundary pixel.
- Compute the similarity between neighbors and connected models.
- Re-label the pixel if the label of the nearest model votes a different label.

Experimental results

Some results on types of sequences and compared with existing methods.



(a) Our method (b) LBP/C (c) LBP-TOP (d) Method in [6] (e) Method in [7]

[6] G. Doretto, A. Chiuso, Y. N. Wu and S. Soatto, Dynamic Texture Segmentation, *ICCV*, 2003
[7] A. Ghoreyshi and R. Vidal, Segmenting Dynamic Textures with Ising Descriptors, ARX Models and Level Sets, *ECCV*, 2006

Experimental results

- Results on sequences *ocean-fire-small*

(a) Frame 8 (b) Frame 21 (c) Frame 40

(d) Frame 60 (e) Frame 80 (f) Frame 100

Experimental results

Chen J, Zhao G & Pietikäinen M (2009) An improved local descriptor and threshold learning for unsupervised dynamic texture segmentation. Proc. ICCV Workshop on Machine Learning for Vision-based Motion Analysis.

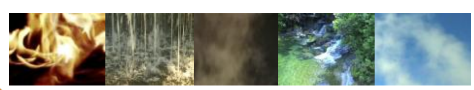
- Results on a real challenging sequence

(a) Frame 5 (b) Frame 10

3.7 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), in press.

- 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*]

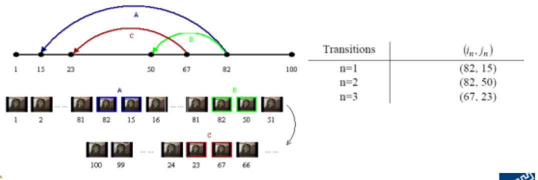
Distance measure can be updated by summing future anticipated costs

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:

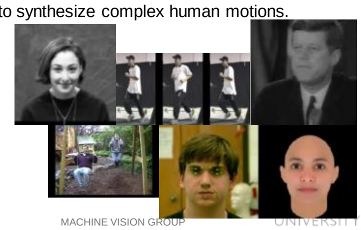


Experiments

- 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. (http://www.ee.oulu.fi/~guoyimo/download/)

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.



Part 4: Summary and future directions

ICCV 2009
September 27, 2009

Matti Pietikäinen
mkp@ee.oulu.fi

Summary

- Modern texture operators form a generic tool for computer vision
- LBP and its spatiotemporal extensions are very effective for various tasks in computer vision
- Spatiotemporal LBP descriptors combine appearance and motion

- The advantages of the LBP methods include
  - computationally very simple
  - can be easily tailored to different types of problems
  - robust to illumination variations
  - robust to localization errors

- For a bibliography of LBP-related research, see http://www.ee.oulu.fi/research/imag/texture

Examples of using LBP in computer vision problems

- Texture segmentation: Ojala et al., 1999;
- Color-texture segmentation: Chen & Chen, 2002; Nammalwar et al., 2003
- Color-texture classification: Mäenpää & Pietikäinen, 2002, 2004
- Color-texture based indexing: Yao & Chen, 2003; Connah & Finlayson, 2006
- Active contour modeling: Savelonas et al., 2006, 2008
- Description of interest regions: Heikkilä et al., 2009
- Object detection: Zhang et al., 2006
- Human detection: Mu et al., 2008
- Crowd estimation: Ma et al., 2008
- On-line boosting: Grabner & Bishof, 2006
- Object classification: Lisin et al., 2005; Autio, 2006
- Recognition of 3D textured surfaces: Pietikäinen et al., 2003
- Background subtraction: Heikkilä & Pietikäinen, 2006
- Object tracking: Takala & Pietikäinen, 2007; Petrovic et al., 2008
- Recognition of dynamic textures: Zhao & Pietikäinen, 2007
- Segmentation of dynamic textures: Chen et al., 2008
- Recognition of actions/events: Kellokumpu et al., 2008; Ma & Cisar, 2009
- Video texture synthesis: Guo et al. 2009

Examples of using LBP in various applications

- Metal strip inspection: Pietikäinen et al., 1994
- Paper inspection: Turtinen et al., 2003
- Inspection of ceramic tiles: Lopes, 2005; Novak & Hocenski, 2005
- Fabric defect detection: Tajeripur et al., 2008
- Quality grading of painted slates: Ghita et al., 2005
- Content-based retrieval: Liao & Chen, 2002; Takala et al., 2005;
- Classification of underwater images: Marcos et al., 2005; Clement et al., 2005
- Aerial image segmentation: Urdiales et al., 2004
- Segmentation of multispectral remote sensing images: Lucieer et al., 2005
- Intravascular tissue characterization: Pujol & Radeva, 2005
- Cell phenotype classification: Nanni & Lumini, 2008
- Ulcer detection in capsule endoscopy images: Li & Meng, 2009
- Mass false positive reduction in mammography: Llado et al., 2009
- Detecting body parts in X-ray images: Jeanne et al., 2009
- Mobile robot navigation: Hong et al., 2002; Davidson & Hutchinson, 2003
- Steganalysis for stenography: Lafferty & Ahmed, 2004
- Designing aesthetically interesting and informative displays: Fogarty et al., 2001

Examples of using LBP in biometrics

- Ovehead view person recognition: Cohen et al., 2000
- Face recognition: Ahonen et al., 2004; G. Zhang et al., 2004; W. Zhang et al. 2005; Li et al., 2006; Rodriguez & Marcel, 2006; Wolf et al., 2008
- Face preprocessing: Heusch et al., 2006; Park & Kim, 2007
- Face detection: Hadid et al., 2004; Jin et al., 2004
- Facial expression recognition: Feng et al., 2004; Shan et al.. 2005, 2009; Liao et al., 2006; Gritti et al., 2008
- Gender classification: Sun et al., 2006; Lian & Lu, 2006; Hadid & Pietikäinen
- Demographic classification: Yang & Ai, 2007
- Eye localization: Kroon et al., 2009
- Iris recognition: Sun et al., 2005; He et al., 2009
- Fingerprint recognition: Nanni & Lumini, 2008
- Palmprint recognition: Wang et al., 2006; Goh et al., 2008
- Grip-pattern recognition in smart gun: Shang & Veldhuis, 2007
- Blind identification of source cell-phone model: Celiktutan et al., 2008
- Facial expression recognition using facial dynamics: Zhao & Pietikäinen, 2007
- Visual speech recognition: Zhao et al., 2009
- Analysis of facial paralysis: He et al., 2009
- Gait recognition: Kellokumpu et al., 2009

Some future directions

- New local descriptors are emerging, for example:
  - WLD - Weber law descriptor (Chen et al., IEEE TPAMI 2009)
  - LPQ - Local phase quantization (Ojansivu & Heikkilä, ICISP 2008)

- Often a single descriptor is not effective enough
- Multi-scale processing
- Use of complementary descriptors
  - LBP&C, LBP&WLD, Haar&LBP, Gabor&LBP, curvelet&LBP, HOG&LBP
- Combining local with more global descriptors (e.g. LBP & Gabor)
- Combining sparse and dense descriptors

- 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 Springer book: Computer Vision Using Local Binary Patterns

To be published next year

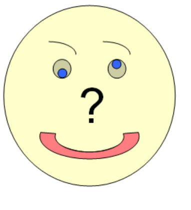
Part I Introduction
1. Background
2. Local binary pattern operators

Part II Analysis of still images
3. Texture classification
4. Segmentation and description of interest regions
5. Applications in image retrieval and 3D recognition

Part III Motion analysis
6. Recognition and segmentation of dynamic textures
7. Background modeling
8. Recognition of actions

Part IV Face analysis
9. Face analysis using still images
10. Face analysis using image sequences
11. Visual speech recognition

Part V Survey to related work
12. Introduction to LBP bibliography



Thanks!