S. Kevin Zhou
Wenyi Zhao
Xiaoou Tang
Shaogang Gong (Eds.)

LNCS 4778

Analysis and Modeling of Faces and Gestures

Third International Workshop, AMFG 2007
Rio de Janeiro, Brazil, October 2007
Proceedings

Springer

S. Kevin Zhou Wenyi Zhao Xiaoou Tang
Shaogang Gong (Eds.)

Analysis and Modeling of Faces and Gestures

Third International Workshop, AMFG 2007
Rio de Janeiro, Brazil, October 20, 2007
Proceedings

Springer

Volume Editors

S. Kevin Zhou
Siemens Corporate Research
755 College Road East, Princeton, NJ 08540, USA
E-mail: kzhou@scr.siemens.com

Wenyi Zhao
Intuitive Surgical Inc.
950 Kifer Road, Sunnyvale, CA 94086, USA
E-mail: wyzhao@ieee.org

Xiaoou Tang
Microsoft Research Asia
5F, Beijing Sigma Center, 49 Zhichun Road, Haidian District, Beijing 100080, China
E-mail: xitang@microsoft.com

Shaogang Gong
Queen Mary, University of London
Department of Computer Science
London E1 4NS, UK
E-mail: sgg@dcs.qmul.ac.uk

Library of Congress Control Number: 2007936691

CR Subject Classification (1998): I.4, I.5, I.3.5, I.2.10, I.2.6, F.2.2

LNCS Sublibrary: SL 6 – Image Processing, Computer Vision, Pattern Recognition, and Graphics

ISSN 0302-9743
ISBN-10 3-540-75689-2 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-75689-7 Springer Berlin Heidelberg New York

Springer is a part of Springer Science+Business Media

springer.com

Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 12174645 06/3180 5 4 3 2 1 0

Lecture Notes in Computer Science 4778

Lecture Notes in Computer Science

Sublibrary 6: Image Processing, Computer Vision, Pattern Recognition, and Graphics

Vol. 4778: S.K. Zhou, W. Zhao, X. Tang, S. Gong (Eds.), Analysis and Modeling of Faces and Gestures. X, 305 pages. 2007.

Vol. 4738: A.C.R. Paiva, R. Prada, R.W. Picard (Eds.), Affective Computing and Intelligent Interaction. XVIII, 781 pages. 2007.

Vol. 4729: F. Mele, G. Ramella, S. Santillo, F. Ventriglia (Eds.), Advances in Brain, Vision, and Artificial Intelligence. XVI, 618 pages. 2007.

Vol. 4713: F.A. Hamprecht, C. Schnörr, B. Jähne (Eds.), Pattern Recognition. XIII, 560 pages. 2007.

Vol. 4679: A.L. Yuille, S.-C. Zhu, D. Cremers, Y. Wang (Eds.), Energy Minimization Methods in Computer Vision and Pattern Recognition. XII, 494 pages. 2007.

Vol. 4678: J. Blanc-Talon, W. Philips, D. Popescu, P. Scheunders (Eds.), Advanced Concepts for Intelligent Vision Systems. XXIII, 1100 pages. 2007.

Vol. 4673: W.G. Kropatsch, M. Kampel, A. Hanbury (Eds.), Computer Analysis of Images and Patterns. XX, 1006 pages. 2007.

Vol. 4642: S.-W. Lee, S.Z. Li (Eds.), Advances in Biometrics. XX, 1216 pages. 2007.

Vol. 4633: M. Kamel, A. Campilho (Eds.), Image Analysis and Recognition. XII, 1312 pages. 2007.

Vol. 4584: N. Karssemeijer, B. Lelieveldt (Eds.), Information Processing in Medical Imaging. XX, 777 pages. 2007.

Vol. 4569: A. Butz, B. Fisher, A. Krüger, P. Olivier, S. Owada (Eds.), Smart Graphics. IX, 237 pages. 2007.

Vol. 4538: F. Escolano, M. Vento (Eds.), Graph-Based Representations in Pattern Recognition. XII, 416 pages. 2007.

Vol. 4522: B.K. Ersbøll, K.S. Pedersen (Eds.), Image Analysis. XVIII, 989 pages. 2007.

Vol. 4485: F. Sgallari, A. Murli, N. Paragios (Eds.), Scale Space and Variational Methods in Computer Vision. XV, 931 pages. 2007.

Vol. 4478: J. Martí, J.M. Benedí, A.M. Mendonça, J. Serrat (Eds.), Pattern Recognition and Image Analysis, Part II. XXVII, 657 pages. 2007.

Vol. 4477: J. Martí, J.M. Benedí, A.M. Mendonça, J. Serrat (Eds.), Pattern Recognition and Image Analysis, Part I. XXVII, 625 pages. 2007.

Vol. 4472: M. Haindl, J. Kittler, F. Roli (Eds.), Multiple Classifier Systems. XI, 524 pages. 2007.

Vol. 4466: F.B. Sachse, G. Seemann (Eds.), Functional Imaging and Modeling of the Heart. XV, 486 pages. 2007.

Vol. 4418: A. Gagalowicz, W. Philips (Eds.), Computer Vision/Computer Graphics Collaboration Techniques. XV, 620 pages. 2007.

Vol. 4417: A. Kerren, A. Ebert, J. Meyer (Eds.), Human-Centered Visualization Environments. XIX, 403 pages. 2007.

Vol. 4391: Y. Stylianou, M. Faundez-Zanuy, A. Esposito (Eds.), Progress in Nonlinear Speech Processing. XII, 269 pages. 2007.

Vol. 4370: P.P. Lévy, B. Le Grand, F. Poulet, M. Soto, L. Darago, L. Toubiana, J.-F. Vibert (Eds.), Pixelization Paradigm. XV, 279 pages. 2007.

Vol. 4358: R. Vidal, A. Heyden, Y. Ma (Eds.), Dynamical Vision. IX, 329 pages. 2007.

Vol. 4338: P.K. Kalra, S. Peleg (Eds.), Computer Vision, Graphics and Image Processing. XV, 965 pages. 2006.

Vol. 4319: L.-W. Chang, W.-N. Lie (Eds.), Advances in Image and Video Technology. XXVI, 1347 pages. 2006.

Vol. 4292: G. Bebis, R. Boyle, B. Parvin, D. Koracin, P. Remagnino, A. Nefian, G. Meenakshisundaram, V. Pascucci, J. Zara, J. Molineros, H. Theisel, T. Malzbender (Eds.), Advances in Visual Computing, Part II. XXXII, 906 pages. 2006.

Vol. 4291: G. Bebis, R. Boyle, B. Parvin, D. Koracin, P. Remagnino, A. Nefian, G. Meenakshisundaram, V. Pascucci, J. Zara, J. Molineros, H. Theisel, T. Malzbender (Eds.), Advances in Visual Computing, Part I. XXXI, 916 pages. 2006.

Vol. 4245: A. Kuba, L.G. Nyúl, K. Palágyi (Eds.), Discrete Geometry for Computer Imagery. XIII, 688 pages. 2006.

Vol. 4241: R.R. Beichel, M. Sonka (Eds.), Computer Vision Approaches to Medical Image Analysis. XI, 262 pages. 2006.

Vol. 4225: J.F. Martínez-Trinidad, J.A. Carrasco Ochoa, J. Kittler (Eds.), Progress in Pattern Recognition, Image Analysis and Applications. XIX, 995 pages. 2006.

Vol. 4191: R. Larsen, M. Nielsen, J. Sporring (Eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI 2006, Part II. XXXVIII, 981 pages. 2006.

Vol. 4190: R. Larsen, M. Nielsen, J. Sporring (Eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI 2006, Part I. XXXVVIII, 949 pages. 2006.

Vol. 4179: J. Blanc-Talon, W. Philips, D. Popescu, P. Scheunders (Eds.), Advanced Concepts for Intelligent Vision Systems. XXIV, 1224 pages. 2006.

Vol. 4174: K. Franke, K.-R. Müller, B. Nickolay, R. Schäfer (Eds.), Pattern Recognition. XX, 773 pages. 2006.

Vol. 4170: J. Ponce, M. Hebert, C. Schmid, A. Zisserman (Eds.), Toward Category-Level Object Recognition. XI, 618 pages. 2006.

Vol. 4153: N. Zheng, X. Jiang, X. Lan (Eds.), Advances in Machine Vision, Image Processing, and Pattern Analysis. XIII, 506 pages. 2006.

Vol. 4142: A. Campilho, M. Kamel (Eds.), Image Analysis and Recognition, Part II. XXVII, 923 pages. 2006.

Vol. 4141: A. Campilho, M. Kamel (Eds.), Image Analysis and Recognition, Part I. XXVIII, 939 pages. 2006.

Vol. 4122: R. Stiefelhagen, J.S. Garofolo (Eds.), Multimodal Technologies for Perception of Humans. XII, 360 pages. 2007.

Vol. 4109: D.-Y. Yeung, J.T. Kwok, A. Fred, F. Roli, D. de Ridder (Eds.), Structural, Syntactic, and Statistical Pattern Recognition. XXI, 939 pages. 2006.

Vol. 4091: G.-Z. Yang, T. Jiang, D. Shen, L. Gu, J. Yang (Eds.), Medical Imaging and Augmented Reality. XIII, 399 pages. 2006.

Vol. 4073: A. Butz, B. Fisher, A. Krüger, P. Olivier (Eds.), Smart Graphics. XI, 263 pages. 2006.

Vol. 4069: F.J. Perales, R.B. Fisher (Eds.), Articulated Motion and Deformable Objects. XV, 526 pages. 2006.

Vol. 4057: J.P.W. Pluim, B. Likar, F.A. Gerritsen (Eds.), Biomedical Image Registration. XII, 324 pages. 2006.

Vol. 4046: S.M. Astley, M. Brady, C. Rose, R. Zwiggelaar (Eds.), Digital Mammography. XVI, 654 pages. 2006.

Vol. 4040: R. Reulke, U. Eckardt, B. Flach, U. Knauer, K. Polthier (Eds.), Combinatorial Image Analysis. XII, 482 pages. 2006.

Vol. 4035: T. Nishita, Q. Peng, H.-P. Seidel (Eds.), Advances in Computer Graphics. XX, 771 pages. 2006.

Vol. 3979: T.S. Huang, N. Sebe, M.S. Lew, V. Pavlović, M. Kölsch, A. Galata, B. Kisačanin (Eds.), Computer Vision in Human-Computer Interaction. XII, 121 pages. 2006.

Vol. 3954: A. Leonardis, H. Bischof, A. Pinz (Eds.), Computer Vision – ECCV 2006, Part IV. XVII, 613 pages. 2006.

Vol. 3953: A. Leonardis, H. Bischof, A. Pinz (Eds.), Computer Vision – ECCV 2006, Part III. XVII, 649 pages. 2006.

Vol. 3952: A. Leonardis, H. Bischof, A. Pinz (Eds.), Computer Vision – ECCV 2006, Part II. XVII, 661 pages. 2006.

Vol. 3951: A. Leonardis, H. Bischof, A. Pinz (Eds.), Computer Vision – ECCV 2006, Part I. XXXV, 639 pages. 2006.

Vol. 3948: H.I. Christensen, H.-H. Nagel (Eds.), Cognitive Vision Systems. VIII, 367 pages. 2006.

Vol. 3926: W. Liu, J. Lladós (Eds.), Graphics Recognition. XII, 428 pages. 2006.

Vol. 3872: H. Bunke, A.L. Spitz (Eds.), Document Analysis Systems VII. XIII, 630 pages. 2006.

Vol. 3852: P.J. Narayanan, S.K. Nayar, H.-Y. Shum (Eds.), Computer Vision – ACCV 2006, Part II. XXXI, 977 pages. 2006.

Vol. 3851: P.J. Narayanan, S.K. Nayar, H.-Y. Shum (Eds.), Computer Vision – ACCV 2006, Part I. XXXI, 973 pages. 2006.

Vol. 3832: D. Zhang, A.K. Jain (Eds.), Advances in Biometrics. XX, 796 pages. 2005.

Vol. 3736: S. Bres, R. Laurini (Eds.), Visual Information and Information Systems. XI, 291 pages. 2006.

Vol. 3667: W.J. MacLean (Ed.), Spatial Coherence for Visual Motion Analysis. IX, 141 pages. 2006.

Vol. 3417: B. Jähne, R. Mester, E. Barth, H. Scharr (Eds.), Complex Motion. X, 235 pages. 2007.

Vol. 2396: T.M. Caelli, A. Amin, R.P.W. Duin, M.S. Kamel, D. de Ridder (Eds.), Structural, Syntactic, and Statistical Pattern Recognition. XVI, 863 pages. 2002.

Vol. 1679: C. Taylor, A. Colchester (Eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI'99. XXI, 1240 pages. 1999.

Preface

The 2007 IEEE International Workshop on Analysis and Modeling of Faces and Gestures (AMFG) is the third workshop of its type organized in conjunction with ICCV, this time in Rio de Janeiro, Brazil. Our primary goal is to bring together researchers and research groups to review the status of recognition, analysis and modeling of face, gesture, activity, and behavior; to discuss the challenges that we are facing; and to explore future directions.

This year we received 55 submissions. Each paper was reviewed by three program committee members. The whole reviewing process was double blind. However, due to size limit, we were only able to accommodate 22 papers, among which 8 are orals and 14 are posters. The topics covered by these accepted papers include feature representation, 3D face, robust recognition under pose and illumination variations, video-based face recognition, learning, facial motion analysis, body pose estimation, and sign recognition.

A special word of thanks goes to Dr. Feng Zhao, our organizing chair, for his dedication and great efforts in maintaining both the online submission system and workshop website and in handling most of the author contacts. We are indebted to the advisory committee members for their valuable suggestions and to the program committee members for their hard work and timely reviews. Finally, we thank Cognitec System GmbH and Siemens Corporate Research for their sponsorship.

October 2007

S. Kevin Zhou
Wen-Yi Zhao
Xiaoou Tang
Shaogang Gong

Organization

AMFG 2007 was held in conjunction with ICCV 2007.

Workshop Chairs

S. Kevin Zhou	Siemens Corporate Research
Wen-Yi Zhao	Intuitive Surgical, Inc.
Xiaoou Tang	Microsoft Research
Shaogang Gong	Queen Mary, University of London

Advisory Committee

Rama Chellappa	University of Maryland
Thomas Huang	University of Illinois at Urbana-Champaign
Anil Jain	Michigan State University

Organizing Chair

Feng Zhao	Chinese University of Hong Kong

Sponsoring Institutions

Cognitec System GmbH
Siemens Corporate Research

Program Committee

Jake Aggarwal	UT Austin, USA
Aaron Bovick	GaTech, USA
Kevin Bowyer	Notre Dame, USA
Rama Chellappa	Maryland, USA
Tsuhan Chen	Carnegie Mellon, USA
Jeff Cohn	Pittsburgh, USA
Robert Collins	Penn State, USA
Tim Cootes	Manchester, UK
James Davis	Ohio State, USA
Larry Davis	Maryland, USA
David Hogg	Leeds, UK

David Jacobs	Maryland, USA
Anil Jain	Michigan State, USA
Mike Jones	MERL, USA
Ron Kimmel	Technion, Israel
Josef Kittler	Surrey, UK
David Kriegman	UCSD, USA
Stan Li	NLPR, China
Chengjun Liu	NJIT, USA
Qingshan Liu	NLPR, China
Jiebo Luo	Kodak, USA
Aleix Martinez	Ohio State, USA
Gerard Medioni	USC, USA
Dimitris Metaxas	Rutgers, USA
Alice O'Toole	UT Dallas, USA
Jonathon Phillips	NIST, USA
Matti Pietikainen	OULU, Finland
Long Quan	UST, HK
Amit Roy-Chowdhury	UCRiverside, USA
Stan Sclaroff	Boston, USA
Matthew Turk	UCSB, USA
Harry Wechsler	George Mason, USA
Shuicheng Yan	UIUC, USA
Ming-Hsuan Yang	Honda Research, USA
Zhengyou Zhang	Microsoft Research, USA
Feng Zhao	CUHK, HK

Author Index

Table of Contents

Oral - I

Poster - I

Oral - II

Poster - II

Learning Personal Specific Facial Dynamics for Face Recognition from Videos

Abdenour Hadid[1], Matti Pietikäinen[1], and Stan Z. Li[2]

[1] Machine Vision Group, P.O. Box 4500, FI-90014, University of Oulu, Finland
http://www.ee.oulu.fi/mvg
[2] Institute of Automation, Chinese Academy of Sciences, 95 Zhongguancun Donglu
Beijing 100080, China

Abstract. In this paper, we present an effective approach for spatiotemporal face recognition from videos using an Extended set of Volume LBP (Local Binary Pattern features) and a boosting scheme. Among the key properties of our approach are: (1) the use of local Extended Volume LBP based spatiotemporal description instead of the holistic representations commonly used in previous works; (2) the selection of only personal specific facial dynamics while discarding the intra-personal temporal information; and (3) the incorporation of the contribution of each local spatiotemporal information. To the best of our knowledge, this is the first work addressing the issue of learning the personal specific facial dynamics for face recognition.

We experimented with three different publicly available video face databases (MoBo, CRIM and Honda/UCSD) and considered five benchmark methods (PCA, LDA, LBP, HMMs and ARMA) for comparison. Our extensive experimental analysis clearly assessed the excellent performance of the proposed approach, significantly outperforming the comparative methods and thus advancing the state-of-the-art.

Keywords: Facial Dynamics, Local Binary Patterns, Face Recognition, Boosting.

1 Introduction

Psychological and neural studies [1] indicate that both fixed facial features and dynamic personal characteristics are useful for recognizing faces. However, despite the usefulness of facial dynamics, most automatic recognition systems use only the static information as it is unclear how the dynamic cue can be integrated and exploited. Thus, most research has limited the scope of the problem by applying methods developed for still images to some selected frames [2]. Only recently have researchers started to truly address the problem of face recognition from video sequences [3,4,5,6,7,8,9].

In [3], an approach exploiting spatiotemporal information is presented. It is based on modeling face dynamics using identity surfaces. Face recognition is performed by matching the face trajectory that is constructed from the discriminating features and pose information of the face with a set of model trajectories constructed on identity surfaces. Experimental results using 12 training sequences and the testing sequences of three subjects were reported with a recognition rate of 93.9%.

S.K. Zhou et al. (Eds.): AMFG 2007, LNCS 4778, pp. 1–15, 2007.

In [4], Li and Chellappa used the trajectories of tracked features to identify persons in video sequences. The features are extracted using Gabor attributes on a regular 2D grid. Using a small database of 19 individuals, the authors reported performance enhancement over the frame to frame matching scheme. In another work, Zhou and Chellappa proposed a generic framework to track and recognize faces simultaneously by adding an identification variable to the state vector in the sequential important sampling method [5].

An alternative to model the temporal structures is the use of the condensation algorithm. This algorithm has been successfully applied for tracking and recognizing multiple spatiotemporal features. Recently, it was extended to video based face recognition problems [6,5]. More recently, the Auto-Regressive and Moving Average (ARMA) model [10] was adopted to model a moving face as a linear dynamical system and perform recognition [7].

Perhaps, the most popular approach to model temporal and spatial information is based on the Hidden Markov models (HMM) which have also been applied to face recognition from videos [8]. The idea is simple: in the training phase, an HMM is created to learn both the statistics and temporal dynamics of each individual. During the recognition process, the temporal characteristics of the face sequence are analyzed over time by the HMM corresponding to each subject. The likelihood scores provided by the HMMs are compared. The highest score provides the identity of a face in the video sequence.

Unfortunately, most of the methods described above use spatiotemporal representations that suffer from at least one of the following drawbacks: (1) the local information which is shown to be important to facial image analysis [11] is not well exploited with holistic methods such as HMMs; (2) while only personal specific facial dynamics are useful for discriminating between different persons, the intra-personal temporal information which is related to facial expression and emotions is also encoded and used; and (3) equal weights are given to the spatiotemporal features despite the fact that some of the features contribute to recognition more than others. To overcome these limitations, we propose an effective approach for face recognition from videos that uses local spatiotemporal features and selects only the useful facial dynamics needed for recognition. The idea consists of looking at a face sequence as a selected set of volumes (or rectangular prisms) from which we extract local histograms of Extended Volume Local Binary Pattern (EVLBP) code occurrences. Our choice of adopting LBP (Local Binary Patterns) for spatiotemporal representation is motivated by the recent results of LBP approach [12] in facial image analysis [13] and also in dynamic texture recognition [14].

In this paper, noticing the limitations of volume LBP operator in handling the temporal information, we first extend the operator and derive a rich set of volume LBP features denoted EVLBP. Then, instead of ignoring the weight of each feature or simply concatenating the local EVLBP histograms computed at predefined locations, we propose an effective approach for automatically determining the optimal size and locations of the local rectangular prisms (volumes) from which EVLBP features should be computed. More importantly, we select only the most discriminative spatiotemporal EVLBP features for face recognition while discard the features which may hinder the recognition process. For this purpose, we use AdaBoost learning technique [15] which has shown its efficiency in feature selection task. The goal is to classify the EVLBP based spatiotemporal features into intra and extra classes, and then use only the extra-class information for recognition. To the best of our knowledge, this is the first work addressing the issue of learning personal specific facial dynamics for face recognition.

2 Extended Volume LBP Features (EVLBP)

The LBP texture analysis operator, introduced by Ojala *et al.* [16,12], is defined as a gray-scale invariant texture measure, derived from a general definition of texture in a local neighborhood. It is a powerful means of texture description and among its properties in real-world applications are its discriminative power, computational simplicity and tolerance against monotonic gray-scale changes.

The original LBP operator forms labels for the image pixels by thresholding the 3×3 neighborhood of each pixel with the center value and considering the result as a binary number. Fig. 1 shows an example of an LBP calculation. The histogram of these $2^8 = 256$ different labels can then be used as a texture descriptor. Each bin (LBP code) can be regarded as a micro-texton. Local primitives which are codified by these bins include different types of curved edges, spots, flat areas etc.

The calculation of the LBP codes can be easily done in a single scan through the image. The value of the LBP code of a pixel (x_c, y_c) is given by:

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

where g_c corresponds to the gray value of the center pixel (x_c, y_c), g_p refers to gray values of P equally spaced pixels on a cicrle of radius R, and s defines a thresholding function as follows:

$$s(x) = \begin{cases} 1, if\ x \geq 0; \\ 0, otherwise. \end{cases} \tag{2}$$

The occurrences of the LBP codes in the image are collected into a histogram. The classification is then performed by computing histogram similarities. For an efficient representation, facial images are first divided into several local regions from which LBP histograms are extracted and concatenated into an enhanced feature histogram. In such a description, the face is represented in three different levels of locality: the LBP labels for the histogram contain information about the patterns on a pixel-level, the labels

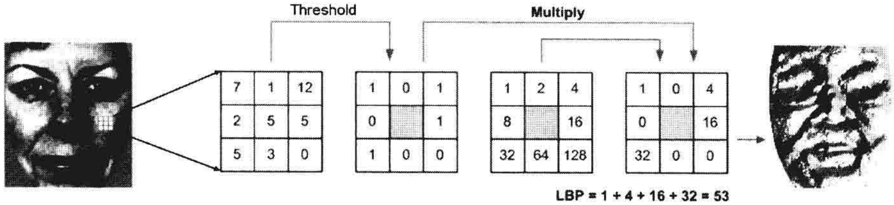

Fig. 1. Example of an LBP calculation

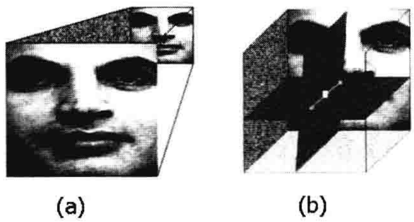

Fig. 2. (a): A face sequence is seen as a rectangular prism and (b): An example of 3D neighborhood of a pixel in Volume LBP

are summed over a small region to produce information on a regional level and the regional histograms are concatenated to build a global description of the face. This locality property, in addition to the computational simplicity and tolerance against illumination changes, are behind the success of LBP approach for facial image analysis [13].

The original LBP operator (and also its later extension to use neighborhoods of different sizes [12]) was defined to deal only with the spatial information. For spatiotemporal representation, Volume LBP operator (VLBP) has been recently introduced in [14]. The idea behind VLBP is very simple. It consists of looking at a face sequence as rectangular prism (or volume) and defining the neighborhood of each pixel in three dimensional space. Fig. 2 explains the principle of rectangular prism and shows an example of 3D neighborhood for Volume LBP.

There are several ways of defining the neighboring pixels in VLBP. In [14], P equally spaced pixels on a circle of radius R in the frame t, and $P+1$ pixels in the previous and posterior neighboring frames with time interval L were used. This yielded in VLBP operator denoted $\text{VLBP}_{L,P,R}$. Fig. 3 (top) illustrates an example of VLBP operator with P=4 and R=1.

We noticed in our experiments on face recognition from videos that $\text{VLBP}_{L,P,R}$ does not encode well enough the temporal information in the face sequences since the operator considers neighboring points only from three frames and therefore the information in the frames with time variance less than L are missed out. In addition, a fixed number of neighboring points (i.e. P) are taken from each of the three frames, yielding in a less flexible operator with large set of neighboring points. To overcome these limitations, we introduce here an extended set of VLBP patterns by considering P points in $frame_{\ t}$, Q points in the $frames_{\ t\pm L}$ and S points in the $frames_{\ t\pm 2L}$. This yields in Extended Volume LBP (EVLBP) operator that we denote by $\text{EVLBP}_{L,(P,Q,S),R}$.

By setting

$$\begin{cases} Q = P+1 \\ S = 0 \end{cases} \tag{3}$$

$\text{EVLBP}_{L,(P,Q,S),R}$ will be equivalent to $\text{VLBP}_{L,P,R}$. Therefore, $\text{VLBP}_{L,P,R}$ can be seen as a special case of $\text{EVLBP}_{L,(P,Q,S),R}$. Fig. 3 (bottom) illustrates an example of Extended Volume LBP operator with P=4, Q=S=1 and R=1 ($\text{EVLBP}_{L,(4,1,1),1}$), while Fig. 3 (top) illustrates an example of $\text{VLBP}_{L,4,1}$ operator which is equivalent to $\text{EVLBP}_{L,(4,5,0),1}$.

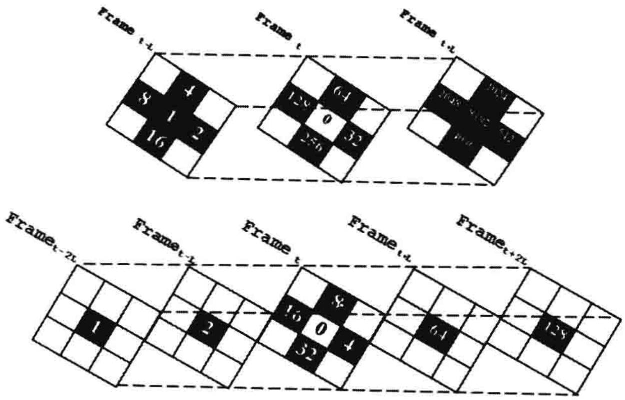

Fig. 3. Top: $VLBP_{L,4,1}$. Bottom: $EVLBP_{L,(4,1,1),1}$

Once the neighborhood function is defined, we divide each face sequence into several overlapping rectangular prisms of different sizes, from which we extract local histograms of EVLBP code occurrences. Then, instead of simply concatenating the local histograms into a single histogram, we use AdaBoost learning algorithm for automatically determining the optimal size and locations of the local rectangular prisms, and more importantly for selecting the most discriminative EVLBP patterns for face recognition while discarding the features which may hinder the recognition process.

3 Learning EVLBP Features for Face Recognition

To tackle the problem of selecting only the spatiotemporal information which is useful for recognition while discarding the information related to facial expressions and emotions, we adopt AdaBoost learning technique [15] which has shown its efficiency in feature selection tasks. The idea is to separate the facial information into intra and extra classes, and then use only the extra-class EVLBP features for recognition.

First, we segment the training face sequences into several overlapping shots of F frames each in order to increase the number of training data. Then, we consider all combinations of face sequence pairs for the intra and extra classes. From each pair ($sequence_i^1$, $sequence_i^2$), we scan both face sequences with rectangular prisms of different sizes. At each stage, we extract the EVLBP histograms from the local rectangular prisms and compute the χ^2 (Chi-square) distances between the two local histograms. χ^2 dissimilarity metric for comparing a target histogram ξ to a model histogram ψ is defined by:

$$\chi^2(\xi, \psi) = \sum_{j=0}^{l-1} \frac{(\xi_j - \psi_j)^2}{\xi_j + \psi_j}, \tag{4}$$

where l is the length of feature vector used to represent the local rectangular prisms.

Thus, for each pair of face sequences, we obtain a feature vector X_i whose elements are χ^2 distances. Let us denote $Y_i \in \{+1, -1\}$ the class label of X_i where $Y_i = +1$