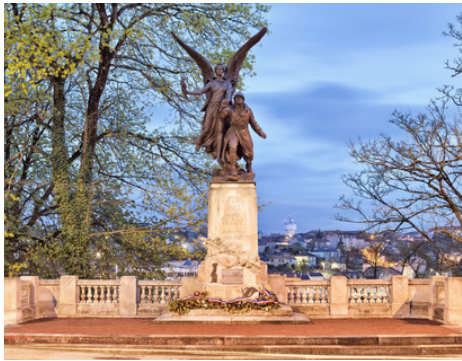
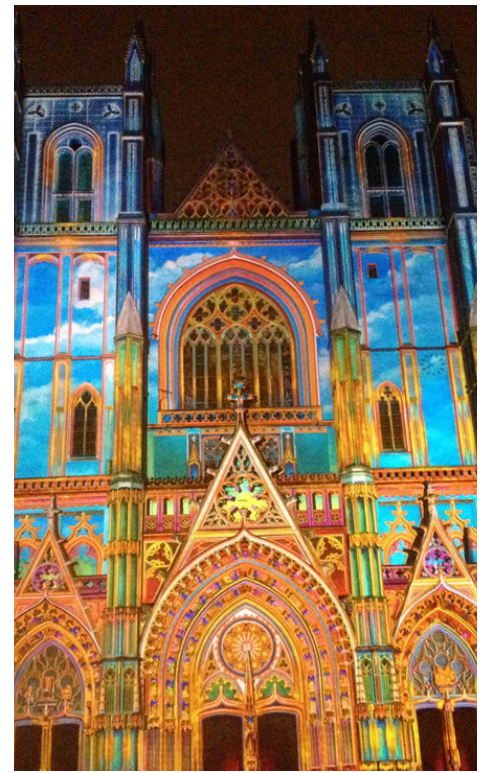


CORESA 2018

COMPRESSION ET
REPRÉSENTATION DES
SIGNAUX AUDIOVISUELS

12-14
NOVEMBRE

Espace Mendes France
Poitiers, France

@CORESA2018

La 20e édition du colloque CORESA 2018 (COmpression et REprésentation des Signaux Audiovisuels) est organisée par le laboratoire XLIM (UMR CNRS 7252) et se déroule à l'Espace Mendes France du 12 au 14 novembre 2018. Le colloque est labellisé par le GdR-ISIS et le Pré-Gdr Sécurité Informatique du CNRS.

Ces journées offrent aux chercheurs et aux praticiens du domaine du multimédia des sessions plénières scientifiques et techniques, des présentations orales, des posters, des démonstrations et des discussions sur des questions pertinentes et stimulantes concernant le futur du multimédia. L'édition de cette année met l'accent sur les nouvelles tendances en compression des signaux visuels, l'imagerie plénoptique et holographique, la réalité virtuelle et augmentée, ...

COMITES

Comité de pilotage

- M. Antonini (I3S, Université de Nice Sophia-Antipolis, CNRS)
- C. Charrier (GREYC, Université de Caen Normandie)
- F. Dupont (LIRIS, Université Claude Bernard Lyon 1)
- J. Jung (Orange Labs)
- L. Morin (IETR, Université de Rennes)
- J-M. Moureaux (CRAN, TELECOM Nancy)
- W. Puech (LIRMM, Université Montpellier)

Comité local d'organisation

- C. Larabi (XLIM, Université de Poitiers) - Responsable

Membres

- B. Bringier (XLIM, Université de Poitiers)
- Y. Fan (XLIM, Université de Poitiers)
- D. Helbert (XLIM, Université de Poitiers)
- S. Jaballah (XLIM, Université de Poitiers)
- B. Mercier (XLIM, Université de Poitiers)
- B. Ray (XLIM, Université de Poitiers)
- T. Urruty (XLIM, Université de Poitiers)

Comité de Programme

M. Antonini	A. Chetouani	E.M. Mouaddib
A. Beghdadi	F. Davoine	J-M. Moureaux
J. Benois-Pineau	F. Dufaux	F. Payan
L. Boubchir	L. Duval	W. Puech
M. Cagnazzo	C. Fernandez-Maloigne	V. Ricordel
A. Caplier	J. Jung	C. Rosenberger
C. Cavaro-Menard	M.C. Larabi	J-P. Vandeborre
M. Chaumont	A. Mokraoui	B. Vozel
C. Charrier	L. Morin	

PROGRAMME CORESA2018

lundi 12 novembre 2018 (TUTORIELS)

09:30 - 11:00	Tatouage et Stéganalyse : approche par ondelettes (Confluence)
11:00 - 11:20	Pause café
11:20 - 12:50	Les normes de codage vidéo HEVC et VVC : tutoriel et état des lieux (Confluence)
12:50 - 14:00	Déjeuner
14:00 - 15:30	Deep Learning and Neural Networks for computer vision and signal processing (Confluence)
15:30 - 15:50	Pause café
15:50 - 17:20	A guided tour of computational modelling of visual attention (Confluence)

mardi 13 novembre 2018

08:30 - 09:00	Inscriptions - Récupération du package CORESA2018	
08:45 - 09:00	Discours d'ouverture (Planétarium) - Organisateur	
09:00 - 10:20	<u>Nouvelles tendances en imagerie (Planétarium)</u>	
09:00 - 09:20	Compression adaptative d'hologrammes numériques en utilisant la technique du "Matching Pursuit" <i>Anas El Rhammad, Patrick Gioia, Antonin Gilles, Marco Cagnazzo</i>	11
09:20 - 09:40	Approximation faible rang basée super-rayons pour la compression de champs de lumières <i>Elian Dib, Mikaël Le Pendu, Xiaoran Jiang, Christine Guillemot</i>	14
09:40 - 10:00	› Adapting HDR images using uniform color space for SDR quality metrics <i>Maxime Rousselot, Olivier Le Meur, Remi Cozot, Xavier Ducloux, Eric Auffret</i>	16
10:00 - 10:20	› Achieving 6 degrees-of-freedom (6DoF) using divergent cameras - <i>Bappaditya Ray, Joël Jung, Mohamed-Chaker Larabi</i>	18
10:20 - 10:40	Pause café	

10:40 - 11:40	Plénière : 3DoF+ and 6DoF immersive activities in MPEG standardization (Planétarium) Joël Jung	
11:40 - 12:40	<u>Biométrie (Planétarium)</u>	
11:40 - 12:00	› Reconnaissance palmaire multi-spectrale basée sur le filtre log-Gabor à multi-résolution et le codage compétitif <i>Yassir Aberni, Larbi Boubchir, Boubaker Daachi</i>	21
12:00 - 12:20	› Iris Recognition via the use of Binarized Statistical Image Features and Hilbert Transform <i>Nabil Hezil and Abdelhani Boukrouche</i>	24
12:20 - 12:40	› Performance of no-reference image quality metric for face images-based biometric systems <i>Xinwei Liu, Christophe Charrier, Marius Pedersen, Patrick Bours</i>	27
12:40 - 14:10	Déjeuner	
14:10 - 15:50	<u>Compression des données multimédia (Planétarium)</u>	
14:10 - 14:30	› Joint optimization of lifting operators for stereo image coding <i>Ismail Bezzine, Mounir Kaaniche, Noor Al-Maadeed, Saadi Boudjit, Azeddine Beghdadi</i>	29
14:30 - 14:50	› Reference list optimization for improvement of compression of 2D-multiview content for free-navigation applications. <i>Pavel Nikitin, Marco Cagnazzo, Joël Jung</i>	32
14:50 - 15:10	› Décomposition multi-échelles de maillages 3D hexaédriques dans le domaine des géosciences. Étude des performances en compression sans pertes. <i>Lauriane Bouard, Laurent Duval, Frédéric Payan, Marc Antonini</i>	34
15:10 - 15:30	› Tree-Structured Point-Lattice Vector Quantization for 3-D Point Cloud Geometry Compression <i>Amira FILALI, Vincent Ricordel, Nicolas Normand</i>	36
15:30 - 15:50	› DNA coding for image storage using image compression techniques <i>Melpomeni Dimopoulou, Marc Antonini, Pascal Barbry, Raja Appuswamy</i>	39
15:50 - 16:10	Pause café	
16:10 - 17:30	<u>Session Poster</u>	
	› A variant of Local Binary Pattern applicated on ear recognition in unconstrained condition <i>Zineb Youbi, Larbi Boubchir, Abdelhani Boukrouche</i>	41
	› Algorithme d'appariement de blocs basé sur l'optimisation de la compensation de disparité pour le codage d'image stéréoscopique <i>Imen Kadri, Gabriel Dauphin, Anissa Mokraoui</i>	44
	› Clustering Algorithm Based on Shuffled Frog Leaping Algorithm <i>Sami Jaballah, Mohamed-Chaker Larabi</i>	46

	› Robust Image Segmentation for Early Plant Diseases Detection on Leaf Arnaud Ahouandjinou, Cina Motamed	49
	› Segmentation-based Animal detection in aerial images <i>Yunfei Fang, Shengzhi Du, Boubaker Daachi, Larbi Boubchir, Karim Djouani</i>	52
18:30 - 23:00	Diner de Gala (Château du Clos de la Ribaudière)	

mercredi 14 novembre 2018

08:30 - 09:00	Inscriptions - Récupération du package CORESA2018	
09:00 - 10:20	<u>Sécurité (Planétarium)</u>	
09:00 - 09:20	› Détection des attaques de présentation en reconnaissance faciale en combinant l'apprentissage profond et les motifs binaires locaux <i>Tarek Ben Charrada, Wael Elloumi, Aladine Chetouani</i>	55
09:20 - 09:40	› Amélioration de la robustesse de la détection d'images altérées pour la compression <i>Boubacar Diallo, Thierry Urruty, Pascal Bourdon, Christine Fernandez-Maloigne</i>	57
09:40 - 10:00	› Etude de l'apport de la reconstruction des régions occultées du visage pour la reconnaissance des expressions <i>Delphine Poux, Benjamin Allaert, Nacim Ihaddadene, Ioan Marius Bilasco, Chaabane Djeraba</i>	59
10:00 - 10:20	› Protection de la vie privée par partage hybride de photos sur les réseaux sociaux <i>Sébastien Beugnon, Pauline Puteaux, William Puech</i>	62
10:20 - 10:40	Pause café	
10:40 - 11:40	Plénière : Light field image processing (Planétarium) Christine Guillemot	
11:40 - 12:40	<u>Analyse de vidéos (Planétarium)</u>	
11:40 - 12:00	› Etude de l'influence du contenu vidéo dans une transmission Pseudo-Analogique de type SoftCast <i>Anthony Trioux, François-Xavier Coudoux, Patrick Corlay, Mohamed Gharbi</i>	65
12:00 - 12:20	› Étude du mouvement d'un motif à l'aide d'un réseau de neurones impulsionnels <i>Veïs Oudjail, Jean Martinet</i>	68
12:20 - 12:40	› Video summarization based on human action detection and recognition <i>Omar Elharrouss, Noor Al-Maadeed, Somaya Al-Maadeed, Ahmed Bouridane, Azeddine Beghdadi</i>	71
12:40 - 14:10	Déjeuner	

14:10 - 14:50	<u>Qualité (Planétarium)</u>	
14:10 - 14:30	› Estimation de la qualité biométrique d'échantillon de voix dans le cadre de la reconnaissance du locuteur <i>Antoine Cabana, Christophe Charrier</i>	73
14:30 - 14:50	› Blind stereopair quality evaluation based on statistics of binocular contrast features <i>YU FAN, Mohamed-Chaker Larabi, Faouzi Alaya Cheikh, Fernandez-Maloigne Christine</i>	76
14:50 - 15:30	<u>Imagerie médicale (Planétarium)</u>	
14:50 - 15:10	› Une méthode originale basée sur l'IRM pour évaluer le phénotype infiltrant des gliomes de bas grade <i>Cyril Brzenczek, Fabien Rech, Yann Gaudeau, Sophie Mezieres, Tiphaine Obara, Marie Blonski, Luc Taillandier, Jean-Marie Moureaux</i>	79
15:10 - 15:30	› Image Registration for Biomedical Images on Irregular Grids <i>Hong Nhung Pham, David Helbert, Pascal Bourdon, Philippe Carre</i>	81
15:30 - 15:50	Pause café	
15:50 - 17:00	<u>Session Poster</u>	
	› Classification des signaux EEG pour BCI à base de descripteurs de Fourier et réseaux de neurones <i>Kathia CHENANE, Youcef Touati, Larbi Boubchir, Boubaker Daachi</i>	84
	› Compression d'images couleur par classification adaptative d'histogrammes en fonction de la densité informationnelle <i>Alain Clément, Benoit Landry</i>	87
	› Design of optimal wavelet image compression scheme based on spline transform and improved coding algorithm <i>Rania Boujelbene, Yousra Ben Jemaa, Larbi Boubchir</i>	90
	› Feature extraction for face recognition based on SSA-NIG and statistical features <i>Basma Ammour, Larbi Boubchir, Toufik Bouden, Messaoud Ramdani</i>	93
	› Tree density estimator using scale-space blobs and saddle features <i>Abril Fernanda Garcia Ramirez, Benoit Tremblais, Christine Fernandez-Maloigne, Matthias Rös</i>	96

3DoF+ and 6DoF immersive activities in MPEG standardization

360 video has recently gained a lot of interest because of the immersive feeling brought by this technology. Unfortunately, the quality of experience is currently not at the expected level. This presentation will recap the main current issues, and report how MPEG-I (Immersive part of MPEG) is tackling them through one short term project, 3DoF+, by adding motion parallax for restricted motion, and a longer term 6DoF project, that releases most of the constraints, in order to achieve free navigation in typical immersive applications.

Biographie :



Joël Jung received the Ph.D. degree in Electrical Engineering from the University of Nice-Sophia Antipolis, Nice, France, in 2000. From 1996 to 2000, he was with the I3S/CNRS Laboratory, University of Nice-Sophia Antipolis, involved in the improvement of video decoders based on the correction of compression and transmission artifacts. He joined Philips Research France in 2000 as a Research Scientist in video coding, post-processing, perceptual models, objective quality metrics, and low-power codecs. He is currently with Orange Labs Paris, and has contributed to next generation 2D image and video coding, HEVC (JCT-VC group), and to 3D multi-view and depth coding (JCT-3V). His current research interests include representation format and compression for Virtual Reality and Augmented Reality, for next generation immersive 6 degrees of freedom video services. It includes immersive communications, and immersive free navigation in user generated and professional contents, considering light-field based techniques. He holds more than 50 patents in the field of video coding.

Light field image processing

Light field imaging is becoming increasingly popular thanks to recent advances in acquisition devices. By capturing light rays emitted along different directions, light fields yield a rich description of the scene, enabling post-capture processing that can be appealing for a variety of applications. However, the huge volume of high-dimensional light field data is an obvious issue for storage, transmission but also for fast processing. Acquisition devices designed so far to capture light fields come with some technological limitations that translate into trade-offs between angular and spatial resolution. This talk will review fundamentals in light field imaging, the related plenoptic function, the main capturing devices and will present fundamental problems in light field image processing.

Plénière soutenue par



Biographie :



Christine Guillemot is currently Director of Research at INRIA (Institut National de Recherche en Informatique et Automatique) in France. She holds a PhD degree from ENST (Ecole Nationale Supérieure des Telecommunications) Paris (1992). From 1985 to 1997, she has been with France Telecom working in the areas of image and video compression for multimedia and digital television. From 1990 to mid 1991, she has worked as visiting scientist at Bellcore Bell Communication research) in the USA. Her research interests are signal and image processing, and in particular 2D and 3D image and video coding, joint source and channel coding for video transmission over the Internet and over wireless networks, and distributed source coding. She has served as Associate Editor for IEEE Trans. on Image Processing (from 2000 to 2003 and from 2014 to 2016), for IEEE Trans. on Circuits and Systems for Video Technology (from 2004 to 2006), IEEE Trans. On Signal Processing (2007-2009). for the Eurasip journal on image communication (2010-2016), and member of the IEEE journal on selected topics in signal processing (2013-2016). She is IEEE fellow since January 2013.

Tutoriel 1: Deep Learning and Neural Networks for computer vision and signal processing

Representation Learning or Deep Learning consists in automatically learning layered and hierarchical representation with various layers abstraction from large amounts of data. This presentation will review the history of the field, the main actors and the major scientific challenges. We will first present a brief introduction into the general challenges in machine learning of high dimensional input, like images, signals and text. We will present common deep models like convolutional neural networks and recurrent networks and various widely used standard tools and problems, like attention mechanisms, transfer learning and learning structured output. Implementing these models in deep learning frameworks (Tensorflow, PyTorch) will be briefly touched. Finally, we will go into more into detail of some selected applications in computer vision and signal processing.

Biographie :



Christian WOLF is associate professor (Maître de Conférences, HDR) at INSA de Lyon and LIRIS, a CNRS laboratory, since sept. 2005. He is interested in computer vision and machine learning, especially the visual analysis of complex scenes in motion: gesture and activity recognition and pose estimation and learning from interactions. His work puts an emphasis on modelling complex interactions of a large amount of variables: Deep Learning, structured models, and graphical models. He received his MSc in computer science from Vienna University of Technology in 2000, and a PhD in computer science from the National Institut of Applied Science (INSA de Lyon), France, in 2003. In 2012 he obtained the habilitation diploma, also from INSA de Lyon. From september 2004 to august 2005 he was assistant professor at the Louis Pasteur University, Strasbourg. Since September 2017 he is on leave (=en délégation) with INRIA, (Chroma group), at the CITI Laboratory

Tutoriel 2: A guided tour of computational modelling of visual attention

Since the first computational model of visual attention, proposed in 1998 by Itti et al. [1], a lot of progress has been made. Progress concern both the modelling in itself and the way we assess the performance of saliency models. Recently, new advances in machine learning, more specifically in deep learning, have brought a new momentum in this field. In this tutorial, we present saliency models as well as the metrics used to assess their performances. In particular, we will empathize new saliency models which are based on convolutional neural networks. We will present different deep architectures and the different loss functions used during the training process. We will conclude this presentation by introducing saccadic models [2,3] which are a generalization of saliency models.

Biographie :



Dr Olivier Le Meur (PhD 2005) is Associate Professor at the University of Rennes 1 (Ecole Supérieure d'Ingénieurs de Rennes). He supervises the research team PERCEPT at IRISA/Rennes. Olivier Le Meur got his Ph.D. from University of Nantes in 2005 and an HDR (Habilitation à Diriger des Recherches) of the University of Rennes 1 in 2014. Before joining the University in 2009, he was a project leader at the research center of Technicolor R&D. He is the author of more than 50 peer-reviewed publications in international conferences and journals in the field of image processing (IEEE TIP, IEEE PAMI, ICIP), computer vision (ECCV, ACCV) and applied perception (Vision Research). His expertise is in the field of image processing, cognitive sciences and computational modelling of visual attention. For more details, please refer to http://people.irisa.fr/Olivier.Le_Meur/ and <http://www-percept.irisa.fr/>

Le standard de codage vidéo HEVC, normalisé en 2013 par l'ISO et l'ITU-T, a permis d'obtenir un gain de 50% en compression par rapport à l'état de l'art. Nous présentons les spécificités du standard HEVC, tant du point de vue des outils de compression que des fonctionnalités, en nous focalisant sur les principales nouveautés par rapport aux normes précédentes. Par ailleurs, l'ISO et l'ITU-T travaillent actuellement à la construction de son successeur, VVC (pour Versatile Video Codec) qui devrait être normalisé fin 2020. Nous présenterons également un état des lieux de ce futur standard qui promet un gain de 50% sur HEVC.

Biographie :



Félix Henry est ingénieur diplômé de Telecom SudParis, Évry, France, et Docteur de Telecom ParisTech, Paris, France. Il a débuté sa carrière dans le domaine du codage d'images fixes en 1995 à Canon Research Center, France, où il a activement participé à la construction du standard JPEG2000. Il a rejoint Orange en 2010, où il a été impliqué dans le développement d'outils de compression pour la norme HEVC. Il participe aujourd'hui à la norme VVC. Ses domaines d'intérêt incluent le codage intra, le codage des coefficients transformés, le parallélisme haut niveau, et les approches en rupture. Félix Henry est l'auteur de plus de 150 brevets dans le domaine du traitement des signaux.

Tutoriel 4 : Tatouage et Stéganalyse : approche par ondelettes

Dans cet exposé, nous proposons de présenter la problématique de la sécurisation des données Multimédia (image, vidéo) à travers à la fois l'enfouissement d'information (tatouage) mais aussi la détection de messages malicieux (stéganalyse). Plus précisément nous ferons un bilan des approches utilisant les qualités des représentations basées ondelettes permettant d'identifier les informations saillantes présentes dans des données Multimédia.

Biographie :



The main objective of my work is the definition and the optimization of multi-scale, multi-resolution for all the color images and videos processing chain, by taking into account the vector type of the information and possibly specificities of the human visual system. Mathematical models essentially studied relate to Space-Scale-Frequency techniques. Space-Scale-Frequency techniques aim at representing information initially present in a nD signal (scalar or vector) in an optimal manner. However, the enlargement of the wavelets transform to higher dimensions than 1 draws some questions on the information description in the transformed space. We have proposed new discrete atomic decompositions for improving the coding of structuring elements of an image or a sequence. In this framework, we have for example studied Quaternionic and Monogenic representation to process color image. New color tools are proposed to analyze the digital spectrum embedded in each of these formalisms and the definition of new approach for Data security (Watermarking, steganalyse).

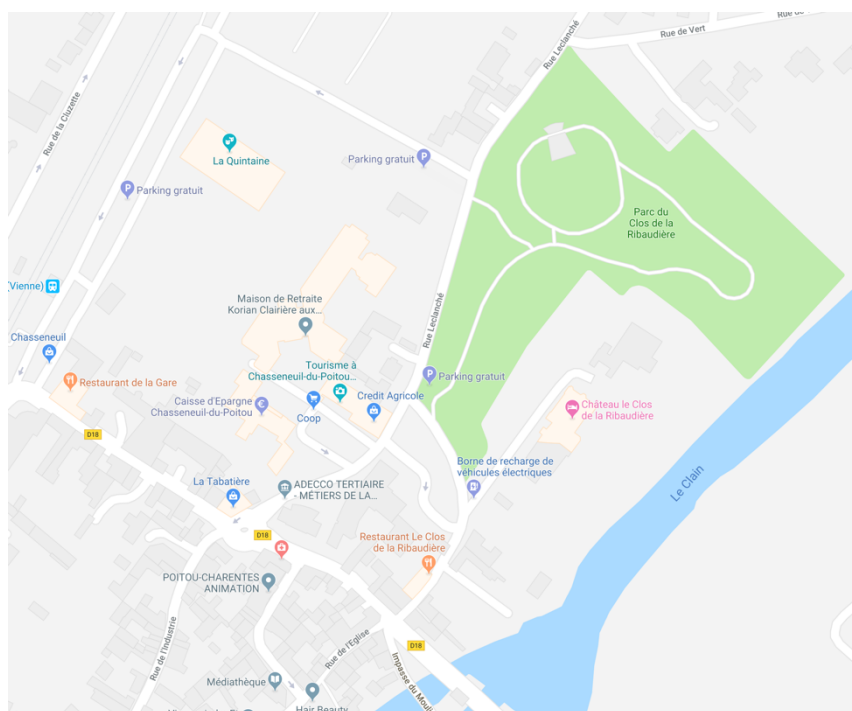
SOIRÉE DE GALA

La soirée de gala a lieu au :



Château du Clos de la Ribaudière
Adresse : 10 Rue du Champ de
Foire, 86360 Chasseneuil-du-Poitou
Tel : 05 49 52 86 66

- A 5 minutes de l'Autoroute A10, de la D910 sur l'axe Poitiers/Châtelleraut.
- A 5 minutes de la Gare du Futuroscope, à moins d'une heure et demi de Paris par le TGV.
- 30 min en Bus (BUS n°1) depuis le centre-ville de Poitiers.



Programme :

18h30 : Départ du bus de l'Espace Mendes France

22h30 : Retour du bus vers le centre-ville de Poitiers.

Veillez à vous munir de la carte du menu que vous avez sélectionné (voir enveloppe badge).

Articles de CORESA 2018



Titre :

Compression adaptative d'hologrammes numériques en utilisant la technique du "Matching Pursuit"

Auteurs :

Anas El Rhammad^(1,3), Patrick Gioia^(1,2), Antonin Gilles⁽¹⁾, Marco Cagnazzo^(1,3)

⁽¹⁾ Institut de Recherche Technologique b<>com, 35510 Cesson-Sévigné, France

⁽²⁾ Orange Labs, 35510 Cesson-Sévigné, France

⁽³⁾ LTCI, Télécom ParisTech, Université Paris-Saclay, 75013, Paris, France

Résumé (100-200 mots) :

L'objectif de cet article est d'étudier la pertinence des ondelettes de Gabor pour la représentation ainsi que la compression des hologrammes numériques en fonction de la position de l'observateur. Le "Matching Pursuit" est utilisé pour une décomposition parcimonieuse des rayons lumineux émis par l'hologramme. Un sous-hologramme est généré en sélectionnant les rayons diffractés vers le champ de visualisation. Dans la littérature, l'utilisation des sous-hologrammes a été proposée comme une alternative à la compression des données massives contenues dans un signal holographique. Notre approche est présentée dans un schéma complet de transmission d'hologrammes en couleur qui peut notamment être utilisé dans le cadre des visiocasques.

Mots-clefs (3 à 5 max.) :

Hologramme numérique, Diffraction, Ondelettes de Gabor, Matching Pursuit, Compression.

Contexte et état de l'art :

L'holographie numérique est l'une des technologies les plus prometteuses pour la visualisation 3D. En effet, elle permet de corriger le conflit entre accommodation et convergence causé par les dispositifs stéréoscopiques classiques, fournissant ainsi à l'utilisateur l'illusion de la profondeur la plus naturelle et réaliste possible. Grâce à ses grandes qualités en termes de confort visuel, l'holographie pourrait trouver application dans le domaine des casques de réalité augmentée ou de réalité virtuelle.

Dans ce contexte, il sera judicieux de représenter le contenu holographique sous un format compressé plus adapté au traitement en temps réel. Étant donné qu'un hologramme numérique est représenté dans un plan 2D, une première approche triviale est d'appliquer les algorithmes de compression classiques utilisés en imagerie. En revanche, un hologramme est un signal très haute fréquence présentant des caractéristiques différentes d'une image classique, limitant les performances de compression obtenues par JPEG2000 ou H.265. D'autres méthodes consistent à décomposer l'hologramme dans une base d'ondelettes plus adaptée aux signaux holographiques (Fresnelets, ondelettes directionnelles...). Malgré l'amélioration apportée par ces techniques, la compression ne permet pas d'avoir une adaptativité en fonction du point de vue, qui consiste à n'encoder que la quantité de données nécessaires pour restituer une perspective donnée.

Étant donné le caractère diffractif des hologrammes, une localisation espace/fréquence précise est indispensable pour l'extraction des directions de diffraction locales. Récemment, nous avons proposé une approche de décomposition parcimonieuse en ondelettes de Gabor basée sur le "Matching Pursuit" [1]. Les différents aspects de compression ont été étudiés et évalués sans prendre en considération le point de vue de l'observateur.

Dans cet article, nous exposons une extension de notre approche dans le cadre d'une application de réalité augmentée où l'utilisateur peut avoir différents points de vue sur la scène. L'idée est d'appliquer une sélection adaptative afin de n'extraire que les rayons lumineux inscrits dans la fenêtre d'observation et ainsi générer le sous-hologramme correspondant à la position de l'utilisateur. Cela permettra de réduire significativement la quantité de données à transmettre à chaque instant sans dégrader la qualité visuelle de la vue reconstruite.

Travail proposé (*Décrire l'objectif du travail et indiquer clairement le problème/défi technique étudié*) :

La localisation spatio-fréquentielle exacte des ondelettes de Gabor leur permettent une décomposition locale de l'hologramme en un ensemble de rayons lumineux. Il se peut que cette expansion ne soit pas optimale en termes de parcimonie en considérant l'hologramme entier, en revanche, elle permet de générer des sous-hologrammes correspondant aux points de vue de l'utilisateur.

La figure 1 présente le cas d'usage de notre approche. Dans un premier temps, un ensemble redondant de coefficients $\{c_n\}_{1 \leq n \leq N}$ est obtenu par le serveur en calculant les produits scalaires entre l'hologramme H et les ondelettes de Gabor complexes du dictionnaire D . Dans un second temps, le "Matching Pursuit" (MP) est appliqué afin de générer un sous-ensemble parcimonieux $\{c_k\}_{1 \leq k \leq K}$, avec ($K \ll N$). Le MP permet d'extraire itérativement les ondelettes les plus corrélées avec H . Étant donnée la position ainsi que l'angle de vue de l'observateur, un module de sélection adaptative (AS) permet l'extraction d'un ensemble $\{c_m\}_{1 \leq m \leq M}$ avec ($M < K$) correspondant aux M coefficients retenus pour une vue V . La correspondance entre les directions d'émission de l'hologramme et les fréquences des ondelettes se fait via l'équation des réseaux [2]

$$\sin \varphi = \lambda f,$$

où φ correspond à l'angle de diffraction du rayon émis, λ est la longueur d'onde et f la fréquence locale.

Par la suite, l'ensemble des coefficients quantifiés $\{[c_n]\}$, leurs positions ainsi que les indices d'ondelettes sont encodés en utilisant un codage arithmétique adaptatif. Le bitstream est ainsi généré et transmis pour décodage au niveau du client. Finalement, le sous-hologramme H_V est calculé et affiché sur un HMD en utilisant un Modulateur Spatial de Lumière (SLM).

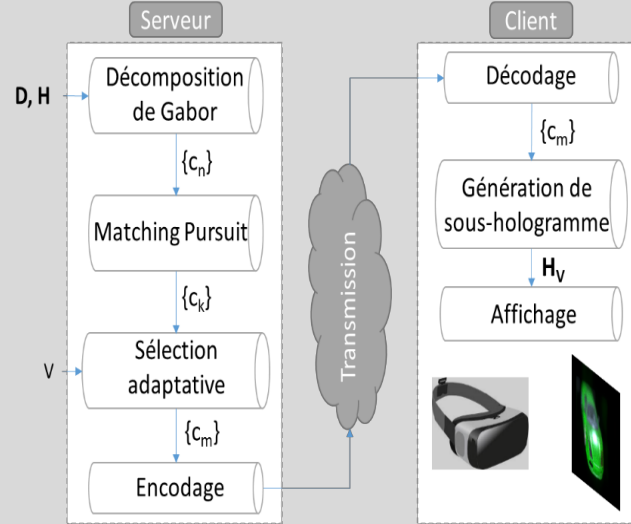


Figure 1 - Transmission adaptative d'un hologramme

Résultats :

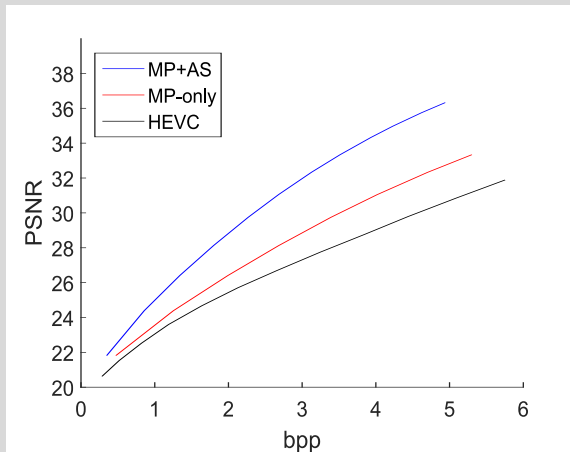


Figure 2 - Courbes RD pour la vue V_{HG}

Afin d'évaluer la performance de notre approche combinant le Matching Pursuit et la sélection adaptative (MP+AS), nous comparons la courbe débit-distorsion (RD) relative à la reconstruction numérique du sous-hologramme H_V avec celles obtenues par encodage de

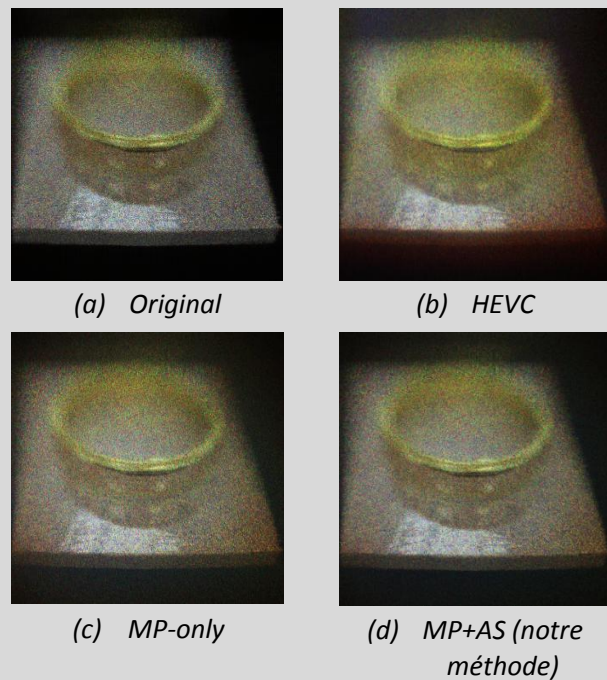


Figure 3 – Reconstructions numériques (2.65 bpp)

l'hologramme H entier en utilisant le Matching Pursuit seulement (*MP-only*) et le codec *HEVC* en mode intra.

La figure 2 illustre les résultats pour un observateur ayant une vue V_{HG} correspondant à une position en haut à gauche vis-à-vis de l'hologramme "Ring". On constate que notre méthode est nettement plus performante que *HEVC* avec un gain moyen de $2.7dB$ calculé sur trois vues différentes. Comparé à *MP-only*, le gain moyen est de $2dB$. La figure 3 présente les reconstructions numériques pour un débit de 2.65 bpp . La meilleure qualité visuelle est obtenue par notre approche. Des résultats de compression détaillés pour d'autres hologrammes et points de vue sont donnés en [3].

Conclusion et perspectives :

Dans cet article nous avons démontré la validité de notre approche destinée à la compression d'hologrammes numériques. Dans un premier temps, la réduction des données à transmettre est obtenue via le Matching Pursuit, ensuite une sélection adaptative sur les ondelettes de Gabor permet d'extraire uniquement les rayons lumineux inscrits dans le champ de vision de l'observateur. Les résultats expérimentaux ont révélé des gains de compression significatifs par rapport aux méthodes de l'état de l'art.

Références (5 max.):

- [1] A. El Rhammad, P. Gioia, A. Gilles, M. Cagnazzo, et B. Pesquet-Popescu. Color digital hologram compression based on matching pursuit. *Applied Optics*, 57(17) :4930–4942, 2018.
- [2] Joseph W. Goodman. *Introduction to Fourier Optics*. Roberts and Company Publishers, Englewood, Colo, 3rd édition, 2005.
- [3] A. El Rhammad, P. Gioia, A. Gilles, M. Cagnazzo, et B. Pesquet-Popescu. View-dependent compression of digital hologram based on matching pursuit. Dans *Optics, Photonics, and Digital Technologies for Imaging Applications V*, volume 10679, page 106790L, 2018.



Titre :

Approximation faible rang basée super-rayons pour la compression de champs de lumière

Auteurs :

Elian Dib, Mikael Le Pendu, Xiaoran Jiang, Christine Guillemot

Résumé (100-200 mots) :

Ce papier décrit une méthode d'approximation faible rang locale basée sur le concept de super-rayons pour la compression de champs de lumière. Les super-rayons peuvent être vus comme un ensemble de super-pixels cohérents sur l'ensemble des vues du champ de lumière. Une méthode d'estimation de disparité est proposée utilisant un a priori faible rang sur les données, ceci afin d'aligner tous les super-pixels formant un super-rayon. Une méthode de construction des super-rayons est décrite permettant de contraindre leur forme et leur dimension, pour gérer les occultations et rendre possible l'approximation faible rang réalisée sur chaque super-rayon. Un algorithme de compression est proposé pour chaque élément du modèle.

Mots-clefs (3 à 5 max.) :

Champs de lumière (light fields), compression, modèles faible rang, sur-segmentation

Contexte et état de l'art :

L'imagerie par champs de lumière offre des perspectives intéressantes pour un grand nombre d'applications de la vision par ordinateur. Mais le très grand volume de données capturées pose de gros problèmes de stockage, d'où la nécessité de développer des schémas de compression efficaces qui exploitent au mieux la forte redondance des données. Dans ce papier, nous nous concentrons sur le problème de la compression de champs de lumière capturés par des caméras plénoptiques utilisant des réseaux de micro-lentilles. La compression de ce type de champs de lumière a connu un fort regain d'intérêt au cours des dernières années. Les méthodes proposées peuvent être classées en deux catégories : les méthodes comprimant les données brut (le capteur 2D) appelées image micro-lentille (ou lenslet) et les méthodes comprimant l'ensemble des vues, ou sous-ouvertures, extraites à partir des données brutes. Les méthodes proposées pour comprimer les images micro-lentilles, pour la plupart, étendent le standard de compression HEVC en y intégrant des modes de prédiction supplémentaires [1]. Les solutions codant les vues, pour la plupart, également étendent ou appliquent les standards existants en codant les vues comme une séquence vidéo [2,3]. En revanche, les auteurs de [4] proposent une approximation faible rang globale et basée homographies pour réduire la dimensionnalité angulaire des données, les paramètres du modèle étant ensuite codés avec HEVC Intra. L'approche proposée ici explore des modèles locaux d'approximation faible rang définis sur des super-rayons. Le concept de super-rayon a été introduit dans [5] comme une extension du concept de super-pixels. Les super-rayons peuvent être vus comme le regroupement des rayons de lumière correspondant à un ensemble de points 3D dans l'espace.

Travail proposé (Décrire l'objectif du travail et indiquer clairement le problème/défi technique étudié) :

L'objectif est de concevoir un schéma de compression de champs de lumière (voir Fig. 1), reposant sur une étape de réduction de la dimensionnalité angulaire des données. L'algorithme proposé s'appuie d'abord sur une nouvelle méthode de construction de super-rayons adaptée à l'objectif d'approximation faible rang. Un super-rayon est constitué d'un super-pixel par vue correspondant au même ensemble de points 3D dans l'espace. L'approximation de chaque super-rayon est d'autant plus précise que les super-pixels formant le super-rayon sont bien alignés. Cela nous a amené à introduire une nouvelle méthode d'estimation de disparité adaptée à ce problème en introduisant un a priori faible rang.

Pour chaque super-rayon, la disparité optimale est estimée en minimisant

$$\arg \min_{d, M} \|SR \circ h^d - M\|_F^2 \text{ s. c. } \text{rang}(M) = r,$$

où SR est la représentation matricielle du super-rayon, chaque colonne i contenant tous les pixels d'une même vue i du super-rayon, M est une matrice de même taille et de rang fixe r , $d = (d_x, d_y)$ contient la

disparité horizontale et verticale du super rayon et h^d est l'ensemble des opérateurs

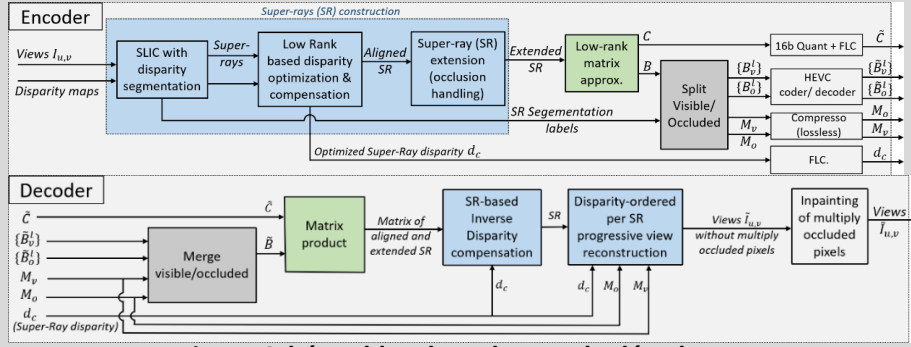


Fig. 1 : Schéma-bloc du codeur et du décodeur

$$h_i^d: (x, y) \mapsto (x + d_x(u_i - u_c), y + d_y(v_i - v_c))$$

appliquant la compensation de disparité d à la vue i . La matrice X contenant dans chaque colonne le super-pixel aligné de chaque vue, sous une forme vectorisée, est alors approximée par le produit d'une matrice faible rang B codée avec HEVC Intra et d'une matrice de coefficients C codée avec un codage fixe sur 32 bits. La matrice B contient, comme illustré en Fig. 2, les composantes principales (ou eigen-images) du modèle.

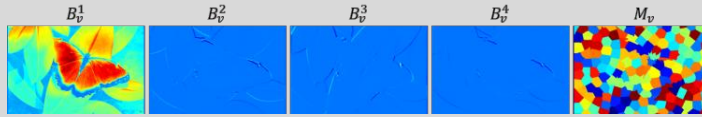


Fig. 2 : Eigenimages (col. 1-4) et segmentation (col.5).

Outre les matrices B et C , il est nécessaire de transmettre au décodeur la carte de segmentation codée en utilisant l'algorithme COMPRESSO et deux valeurs de disparité par super-rayon codées sur 32 bits.

Résultats :

Le tableau 1 présente les gains en PSNR (métrique de Bjontegaard) obtenus avec deux variantes de notre méthode ainsi que HLRA [4] en prenant pour référence HEVC-Lozange [2]. Nous observons un gain important par rapport à [2] sur les données issues de scènes naturelles (ICME, INRIA) et significatif par rapport à [3] sur les données synthétiques (HCI) où la disparité est plus grande.

Ensemble de données	Nom	HLRA [4]	LLRA, zeros	LLRA, récursif
ICME	Fountain_Vincent_2	1.772	0.33518	0.45653
	Friends_1	2.9987	1.8307	1.8672
	Stone_Pillars_Inside	1.9897	1.8486	1.9404
	Vespa	1.7131	0.66977	0.74056
INRIA	Bench	2.8323	1.8657	1.924
	Fruits	2.3055	1.0611	0.94527
	Toys	1.5324	1.2173	1.2851
HCI	Buddha	0.42897	-0.3163	0.60153
	StillLife	-7.6687	-7.6675	-1.6588

Tableau 1 : Comparaison des gains en PSNR par rapport à HEVC-Lozange [2].

Conclusion et perspectives :

Nous avons présenté un nouveau schéma de compression de champs de lumière. Celui-ci s'appuie sur une approximation faible rang locale de super-rayons alignés par compensation de disparité pour exploiter au mieux la réduction de dimensionnalité.

Références (5 max.):

- [1] R. Monteiro, L. Lucas, C. Conti, P. Nunes, N. Rodrigues, S. Faria, C. Pagliari, E. da Silva, and L. Soares, "Light field HEVC-based image coding using locally linear embedding and self-similarity compensated prediction," in IEEE Int. Conf. on Multimedia Expo Workshops (ICMEW), Jul. 2016.
- [2] M. Rizkallah, T. Maugey, C. Yaacoub, and C. Guillemot, "Impact of light field compression on focus stack and extended focus images," in 24th European Signal Processing Conf. (EUSIPCO) Aug. 2016, pp. 898–902.
- [3] D. Liu, L. Wang, L. Li, Z. Xiong, F. Wu, and W. Zeng, "Pseudo-sequence-based light field image compression," in IEEE Int. Conf. on Multimedia Expo Workshops (ICMEW), Jul. 2016.
- [4] X. Jiang, M. Le Pendu, R. Farrugia, and C. Guillemot, "Light field compression with homography-based low-rank approximation," IEEE J. on Selected Topics in Signal Processing, special issue on light field image processing, Oct. 2017.
- [5] M. Hog, N. Sabater, C. Guillemot, Super-rays for efficient light field processing, IEEE J. on Selected Topics in Signal Processing, special issue on light field image processing, Oct. 2017



Adapting HDR images using uniform color space for SDR quality metrics

Auteurs:

Maxime Rousselot, Eric Auffret, Xavier Ducloux
Harmonic inc.

Olivier Le Meur, Rémi Cozot
Univ Rennes, CNRS, IRISA

Résumé (100-200 mots):

Screens capabilities have greatly increased to display images with large ranges of dynamics of luminance in recent years. Because of their new characteristics, the quality assessment of these images has become an active field of studies. Indeed, the metrics created for standard dynamic range (SDR) images are not adapted anymore.

Because most SDR metrics assume that the representation of images is perceptually linear, we propose to study the impact of three uniform color spaces developed specifically for high dynamic range and wide color gamut images, namely, ICtCp [2], J_za_zb_z [3], HDR-Lab [4]. We found that the best color spaces for quality assessment was dependent on the studied database and that the used color space should be chosen in function of the images or video characteristics.

Mots-clefs (3 à 5 max.):

Image quality, Dynamic Range, Image representation

Contexte et état de l'art:

With the progress of screen technology, images can be displayed with a higher dynamic range of luminance. Assessing the quality of these High Dynamic Range (HDR) images is essential to evaluate the performance of a number of image processing such as image compression but SDR quality metrics can't be used directly. Indeed the SDR metrics assumed a luminance range approximately between 1 and 100 cd/m² while HDR images can have a luminance range going below 0.01cd/m² and up to 10000 cd/m².

Existing HDR metrics can be divided in two categories. The first one relies on the modelling of the human visual system (HVS) such as for HDR-VDP2 and HDR-VQM. These metrics are efficient but are quite complex and time-consuming. Another solution is to adapt the SDR quality metrics to HDR images by using an adapted color space. These color spaces ensure that a difference of value always corresponds to the same amount of visually perceived change. In [1], Aydın et al. encoded the luminance using a Perceptual Unit (PU) mapping function. This function presents the same behavior than the classical gamma function for the luminance range corresponding to the SDR displays (between 0.1 to 80 cd/m²) and keeps perceptual linearity above these values.

Recently, other perceptually uniform color spaces which are better to decorrelate chromatic and luminance component were created: ICtCp [2], HDR-Lab [4] (the HDR version of the well-known CIELab color space) and J_za_zb_z [3]. In this paper, we study the impact of using such color spaces for PSNR, SSIM, MS-SSIM and FSIM, four SDR quality metrics.

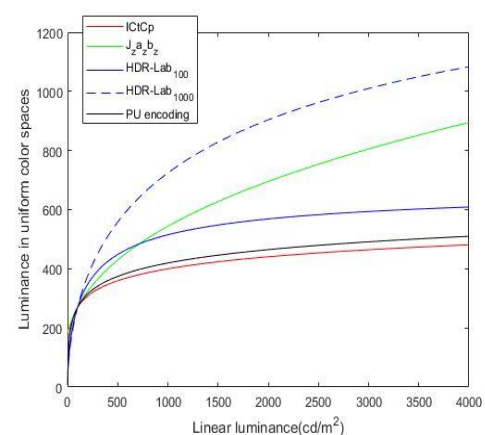


Figure 1: Perceptually linear luminance in function of the linear luminance

Travail proposé (Décrire l'objectif du travail et indiquer clairement le problème/défi technique étudié) :

In this section, we describe our method to adapt the SDR metric and the method used to assess the metrics performance.

First, it is important to tailor the estimation of the perceptually linear luminance for each color space. Indeed, each color space uses its own range of values and some of the considered metric have constants

defined for values between 0 and 255. Then, we scaled the different HDR perceptually linear luminance to fit a similar range than PU between 0 and 100 cd/m² because the PU luminance is already adapted to the SDR metrics. We chose the PU space as a reference because its range is already adapted to the metrics. For the HDRLab color space, two parameters have to be known: the surround luminance and the diffuse white. Because the diffuse white of an image can be difficult to estimate, two values (100 and 1000 cd/m²) are tested. The different rescaled luminance representations can be found on the figure 1.

To assess the performances of the aforementioned metrics, we used three existing HDR databases annotated with subjective scores. To these 3 experiments, the SIM2 HDR display (dynamic range between 0.1 and 4000 cd/m², BT709 gamut) is used. To complete the study, we created two more databases (one HD and one 4K using another display, the Sony BVM-X300 (dynamic range between 0.001 and 1000 cd/m², BT2020 gamut). For the 4K database, we downsampled the images before applying any metric.

Résultats:

Table 1 represents the Spearman rank order correlation coefficient (SROCC) between the subjective and the objective scores of two of the studied databases with four different metrics and five perceptually uniform luminance representations. For the proposed 4K database, the ICTCp, the HDR-Lab100 and the PU encoding space provides the best performances. This suggests that this database have a diffuse white nearer 100 cd/m² than 1000 cd/m². The Narwaria et al. [5] database has more contrasted results. The Jzazbz color space performs better for the SSIM and the MS-SSIM metric. However, for the FSIM metric the PU-encoding yields slightly higher performances. Because the database has images coming from various sources, it is not sure that the diffuse white is homogeneous across the entire database, which could explain the contradictions. On this database, the PSNR gives no convincing performances whatever the used color space. Moreover, for the two databases, we confirm that HDR-VDP2, a more complex metric, does not bring significant additional performance gain than the FSIM and the MS-SSIM metrics used with adapted HDR images.

4K database	PU	ICTCp	Jzazbz	HDR-Lab 100	HDR-Lab 1000
PSNR	0.726	0.742	0.553	0.659	0.467
SSIM	0.707	0.675	0.649	0.697	0.605
MSSSIM	0.851	0.835	0.828	0.849	0.773
FSIM	0.905	0.910	0.885	0.899	0.842
HDR-VDP2	0.8678				
Narwaria et al. [5]	PU	ICTCp	Jzazbz	HDR-Lab 100	HDR-Lab 1000
PSNR	0.533	0.574	0.463	0.470	0.419
SSIM	0.724	0.723	0.772	0.713	0.724
MSSSIM	0.865	0.871	0.909	0.820	0.852
FSIM	0.869	0.790	0.810	0.859	0.840
HDR-VDP2	0.8906				

Table 1: SROCC results for two different databases

Conclusion et perspectives:

In this paper, we analyzed the impact of several perceptually uniform color spaces on several metrics. Using such color spaces can give similar results than using the PU-space or than the more complex metric HDR-VDP2. Using the ICTCp color space can be easier than the PU-space in some contexts: this color space can be used by well-known compression codecs such as HEVC. Moreover, using such color spaces will facilitate the adaptation of quality metrics, which also consider chromatic distortions. Such distortions are more prone to happen when compressing images with a wide color gamut such as BT2020.

However, the result suggests an impact on the quality assessment from the diffuse white (which is often unknown) or to other characteristics often linked to the diffuse white such as the perceived brightness. Further analysis will be needed to fully understand their impacts on quality assessment.

Références (5 max.):

- [1] T. O. Aydin, R. Mantiuk, and H.-P. Seidel, "Extending quality metrics to full luminance range images," in Human Vision and Electronic Imaging XIII, vol. 6806. SPIE, 2008.
- [2] Image parameter values for high dynamic range television for use in production and international programme exchange. Rec BT.2100-1, ITU-R, 2017
- [3] M. Safdar, G. Cui, Y. J. Kim & M. R. Luo, "Perceptually uniform color space for image signals including high dynamic range and wide gamut" in Optics express, 25.13, 2018
- [4] M.D Fairchild, P.H. Chen, "Brightness, lightness, and specifying color in high-dynamic-range scenes and images", in Image Quality and System Performance VIII. Vol. 7867. SPIE, 2011
- [5] M. Narwaria, M.P. Da Silva, P. Le Callet, R. Pepion, "Tone mapping-based high-dynamic-range image compression: study of optimization criterion and perceptual quality" in Optical Engineering Vol.52, 2013



Titre :

Achieving 6 degrees-of-freedom (6DoF) using divergent cameras

Auteurs :

Bappaditya Ray, Joel Jung, Mohamed-Chaker Larabi

Résumé (100-200 mots) :

With the rapid emergence of various 360 video capturing devices and head-mounted displays, 360 videos are becoming of paramount interest for providing immersive experience. The current techniques do not guarantee a good quality of experience mainly because of motion sickness issues. One of the reasons is that they do not take advantage of the parallax between the divergent views, which may be helpful to provide the correct view according to the user's head motion. In this paper, we propose to get rid of the classical ERP projection, and to synthesize arbitrary views using different divergent views together with their corresponding depths. Thus, we can exploit the parallax between the divergent views. In this context, we assess the feasibility of depth estimation and view synthesis using state-of-the-art techniques. Simulation results confirmed the feasibility of such a proposal, in addition to the possibility to achieve sufficient visual quality for a head motion up to 0.1m from the rig, when using generated depth maps for view synthesis.

Mots-clefs (3 à 5 max.) :

360 Video, 6DoF, View Synthesis, Depth Estimation.

Contexte et état de l'art :

360 video uses spherical representation, obtained by stitching the different divergent views, with a mapping on a 2D plane, thanks to projection schemes such as the equirectangular projection (ERP) or the cubemap projection (CMP). The projection only considers two degrees of freedom (DoF), i.e. yaw and pitch (rotations), bringing a relatively low quality of experience. This solution could be seen as a very preliminary step to achieve immersion. MPEG-I currently addresses 3DoF+ activities, aiming to provide parallax information allowing the user to change the viewpoint in a limited space [1]. It is proposed to deal with motion parallax at the system level by extending the MPEG omnidirectional media format (OMAF V1) with metadata, containing depth information, alpha key and patch of residual video [2]. In [3], an omnidirectional video plus depth format (OVD) is proposed using an omnidirectional scene and its corresponding depth to synthesize arbitrary omnidirectional views. Depth is estimated from omnidirectional stereo pairs when the captured depth is not available. These investigations, based on ERP omnidirectional views augmented with auxiliary information or metadata, are expected to bring a significant improvement over the previous 3DoF MPEG-I phase.

The quality of experience is an important issue. For instance, motion sickness results from the conflict between the perceived motion and the expectation of the brain. Current head-mounted devices (HMDs) manage to estimate users head motion precisely. However, the correct view or even more precisely the correct pixels, according to the user's movement, are not displayed, as the current scheme lacks the information of parallax, necessary for correct view rendering. In this contribution, we explore the ability to provide omnidirectional 6DoF by getting rid of the 2D omnidirectional representation. The parallax between different divergent views is used in association with the captured 360 video.

Travail proposé (Décrire l'objectif du travail et indiquer clairement le problème/défi technique étudié) :

It is proposed to use divergent views and the corresponding depths to synthesize arbitrary views. Accordingly, the captured views and the corresponding depth are encoded without any projection, and synthesis of arbitrary views is performed using decoded views and depths. To assess the feasibility of the proposed scheme, we perform the synthesis with original (uncompressed) views and depths. However, in a realistic scenario, the coding of both views and the associated depths should also be involved, as illustrated in Fig. 1.

Two different synthetic contents called WhiteRoom and Kitchen (static scene, single time instance), generated with Blender software are used in the following experiments. For both contents, we arrange eight cameras in a

circular rig equidistantly, as illustrated in Fig. 2. The field of View (FoV) of all the cameras is 90 degrees and the radius of the rig is 0.1m.

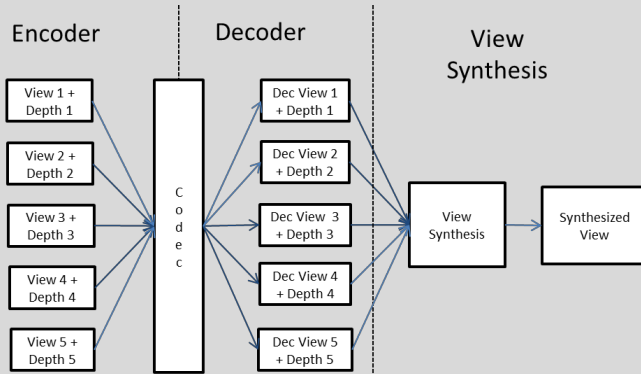


Figure 1 : proposed scheme

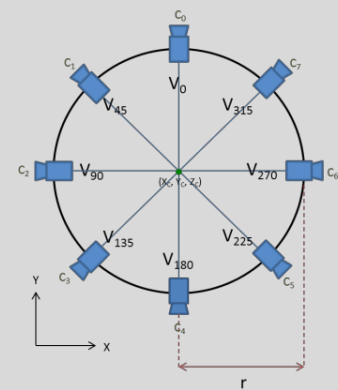


Figure 2 : adopted camera arrangement

Résultats :

By using DERS-6.1, the MPEG depth estimation software [4], to estimate depth, the achieved maps have a very poor quality as shown on Fig. 3.

For the view synthesis, two different scenarios are adopted: step-in/out and move horizontally, as shown in Fig. 4. For both the scenarios, view synthesis is performed for computed and estimated depth with VSRS-4.2, the MPEG view synthesis reference software [5]. From the results, it can be observed from Fig.5 and Fig.6 that the synthesis quality degrades with the amount of movement from the rig. Moreover, the synthesis quality is poor for estimated depths.

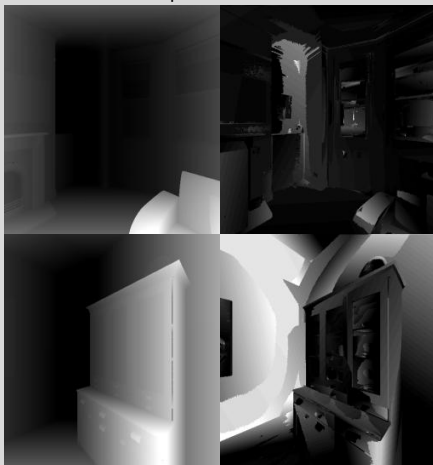


Figure 3 : computed (left) and estimated (right) depth for WhiteRoom (top) and Kitchen (bottom).

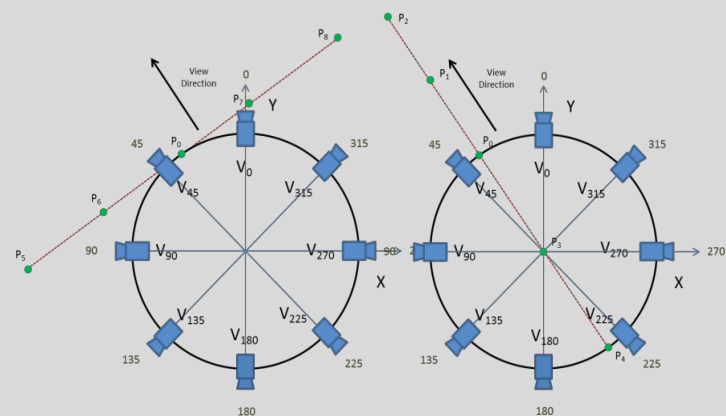


Figure 4 : view synthesis scenario – move horizontally (left), step-in/out (right).

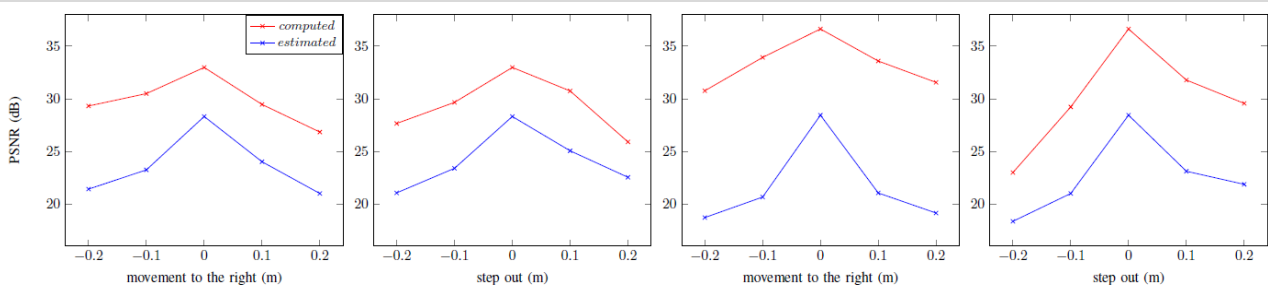
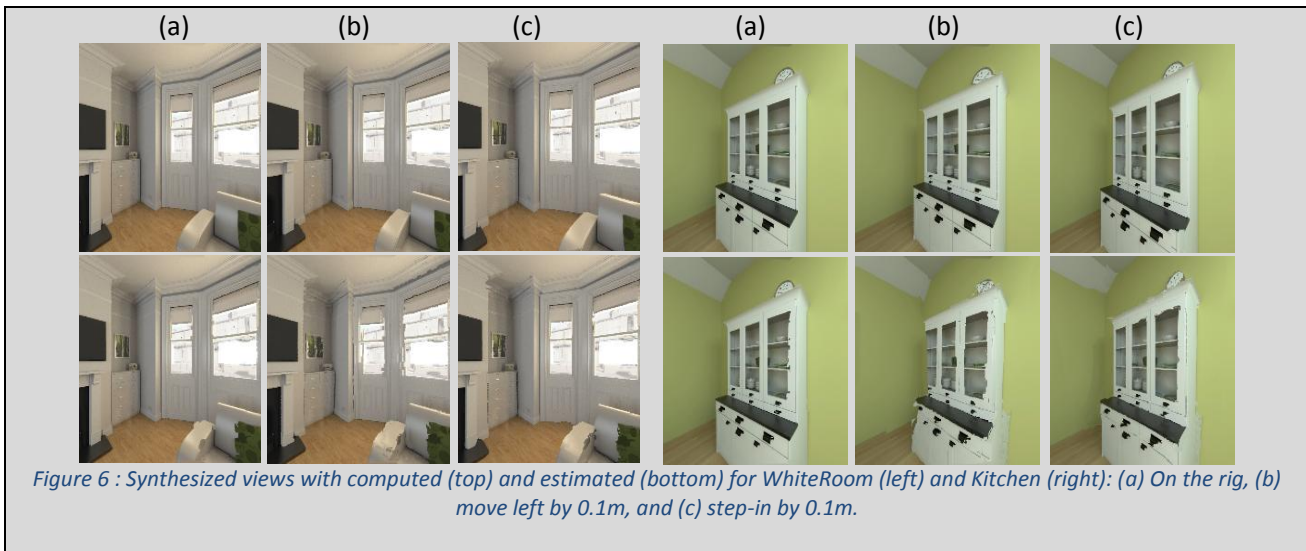


Figure 5 : PSNR of synthesized views for WhiteRoom (left) and Kitchen (right).



Conclusion et perspectives :

We explore a novel way to achieve omnidirectional 6DoF by using the parallax between different divergent views. We show that depth estimation is possible using this parallax, but unfortunately, state-of-the-art depth estimation techniques provide insufficient quality and even fail in the presence of homogeneous areas. View synthesis with the estimated depth provides an acceptable quality when the position of the user is on the camera rig. Any movement out of the rig results in a significant drop in synthesis quality, thus making the synthesized picture visually unacceptable. Synthesis with generated depth provides much better quality, and a movement of 0.1m from the rig can be performed while maintaining an acceptable quality. However, the view synthesis still fails in the presence of reflection and transparent objects.

Those preliminary results were obtained with unmodified VSRS and DERS tools. It is obvious that significant improvements are possible. Furthermore, efficient compression of divergent views should be investigated for a realistic scenario.

Références (5 max.):

- [1] R. Koenen and M. L. Champel, "Requirements MPEG-I phase 1b," in MPEG-requirements of ISO/IEC JTC1/SC29/WG11. MPEG Doc. n17331, 2018, pp. 1–7.
- [2] R. Dore and M. L. Champel, "Information on 3-DoF+ interactive parallax implementation," in MPEG-systems of ISO/IEC JTC1/SC29/WG11. MPEG Doc. m42094, 2018, pp. 1–12.
- [3] K. Wegner, O. Stankiewicz, and M. Domanski, "Considered format for omnidirectional 6-DoF/3-DoF+," in MPEG-I visual of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11. MPEG Doc. m40807, 2017, pp. 1–3.
- [4] K. Wegner, O. Stankiewicz, M. Tanimoto, and M. Domanski, "Enhanced depth estimation reference software (DERS) for freeviewpoint television," in MPEG-FTV of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11. MPEG Doc. m31518, 2014, pp. 1–7.
- [5] K. Wegner, O. Stankiewicz, M. Tanimoto, and M. Domanski, "Enhanced view synthesis reference software (VSRS) for free-viewpoint television," in MPEG-FTV of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11. MPEG Doc. m31520, 2017, pp. 1–4.



Titre :

Reconnaissance palmaire multi-spectrale basée sur le filtre log-Gabor à multi-résolution et le codage compétitif

Auteurs :

Yassir ABERNI, Larbi BOUBCHIR, Boubaker DAACHI
LIASD, Université Paris 8, France

Résumé (100-200 mots) :

La sécurité biométrique devient de plus en plus un outil important pour améliorer la sécurité et apporter une plus grande commodité aux services nécessitant une authentification et une protection des données. Le déploiement des systèmes de sécurité et reconnaissance biométrique à l'aide de traits physiologiques ou comportementaux pour la vérification et l'identification d'individus est d'une grande utilité dans diverses applications. Il s'agit par exemple, du contrôle d'accès, de la surveillance vidéo, et l'expertise médico-légale, ce qui accroît l'intérêt de la recherche pour améliorer ces systèmes et les rendre plus fiables. Plusieurs modalités biométriques ont été étudiées et utilisées pour la reconnaissance biométrique, telles que l'empreinte palmaire. Cette dernière de part son caractère non intrusif, est devenue une modalité attrayante pour sa fiabilité, sa flexibilité et sa stabilité. Bien que de nombreuses méthodes de reconnaissance basées sur l'empreinte palmaire aient été proposées et appliquées avec succès pour l'authentification, la plupart de ces recherches n'utilisent généralement que les images captées à la lumière naturelle. Il est difficile d'améliorer encore la précision de la reconnaissance sur la base de ces images palmaires en raison des limitations de l'utilisation de la lumière naturelle, parmi ces limitations on peut en citer la capacité des attaques par usurpation d'identité, la dégradation des empreintes palmaires avec le temps en raison de plusieurs facteurs tels que l'environnement, l'origine ethnique ou l'âge. L'imagerie multi-spectrale a donc été utilisée pour surmonter ces limitations. En effet, les images palmaires multi-spectrales sont collectées sous différentes bandes spectrales culminant à différentes longueurs d'onde de lumière. Chaque bande spectrale met en évidence des caractéristiques spécifiques de l'empreinte palmaire. Basé sur le principe du codage compétitif, nous avons proposé une méthode de reconnaissance palmaire utilisant : (i) le filtre log-Gabor à multi-résolution permettant de générer une carte de caractéristiques composée des codes gagnants des différentes réponses des filtres les plus bas ; et (ii) la distance de Jaccard pour la mise en correspondance entre les cartes des caractéristiques pour la reconnaissance (i.e. identification et vérification).

Mots-clefs (3 à 5 max.) :

Biométrie multi-spectrale, reconnaissance palmaire, imagerie multi-spectrale, filtre de Gabor 2D, codage compétitif.

Contexte et état de l'art :

Les systèmes d'imagerie multi-spectrale ont été présentés comme une nouvelle technologie capable de détecter des cibles où les images couleur classiques ne pouvaient donner des résultats. Le développement de nouvelles caméras et de nouveaux filtres a permis de voir au-delà du spectre visible, par exemple les ondes infrarouge [700nm-1mm], ultraviolet [10nm-380nm], [0,01nm-10nm]. L'image multi-spectrale rend le processus de modélisation plus précis pour l'objet observé. En appliquant un code de transfert de rayonnement, cette modélisation permet généralement de contraindre la température et la composition de l'objet. L'étape cruciale dans un système de reconnaissance palmaire consiste sur le processus d'extraction de caractéristiques. Ce dernier permet de fournir les informations les plus discriminantes à partir de la région d'intérêt extraite, dite ROI (permet de normaliser la zone de traitement). Pour ce faire, diverses approches de reconnaissance palmaire multi-spectrale ont été proposées dans la littérature. Ces approches peuvent être regroupées en cinq types: (1) les approches structurelles basées principalement sur la géométrie de la main, les lignes principales ou les rides. (2) Les approches basées sur les apparences également appelées approches des sous-espaces, sont des méthodes qui ne nécessitent aucune connaissance préalable de l'empreinte palmaire et associent dans la plupart des cas à des techniques d'analyse de données (e.g. ACP, ACI). (3) Les approches statistiques qui se divisent principalement en approches locales et globales ; où les approches locales transforment l'image en un autre domaine et la divise en petites régions pour extraire les caractéristiques tandis que les approches globales utilisent l'image entière pour l'extraction des caractéristiques. (4) Les approches de type codage basées sur une représentation binaire des réponses des filtres comme les filtres de Gabor et log-Gabor. (5) Les approches hybrides qui peuvent être une combinaison de deux ou plusieurs méthodes.

Travail proposé (Décrire l'objectif du travail et indiquer clairement le problème/défi technique étudié) :

Parmi les différents types d'approches de reconnaissance palmaire multi-spectrale proposées récemment, les approches basées sur le codage compétitif ont prouvé leur efficacité en terme de taux de reconnaissance élevés. Dans le but d'améliorer encore les performances de ces approches, nous avons proposé une nouvelle méthode de reconnaissance palmaire multi-spectrale basée sur le codage compétitif et le filtre log-Gabor multi-échelles.

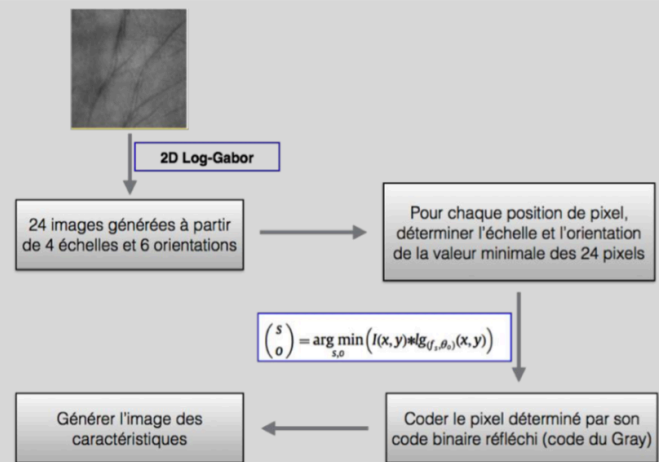
Le choix d'une méthode de classification ou d'extraction des caractéristiques biométrique efficace, peut se traduire par sa capacité de représenter les informations biométriques sous une forme plus discriminante. Le filtre Log-Gabor s'est révélé particulièrement utile dans les applications de traitement d'images, étant donné qu'il permet de mieux capturer les données statistiques des images naturelles. Les données issues de la réponse du filtre Log-Gabor peuvent constituer une base d'informations importantes, car elles peuvent représenter localement des informations de fréquences [6].

Le filtre log-Gabor est par définition la multiplication d'une fonction harmonique et d'une fonction gaussienne modulée par une sinusoïde complexe.

Sa particularité par rapport au filtre Gabor classique est sa réponse en fréquence DC nulle. Bien que le filtre Gabor réalise un sentiment d'optimalité en termes de compromis espace- fréquence, dans certaines applications, il pourrait ne pas être un filtre idéal. Sur certaines largeurs de bande, le filtre Gabor a une composante DC non nulle. Cela signifie que la réponse du filtre dépend de la valeur moyenne du signal. Si la sortie du filtre doit être utilisée pour une application telle que la reconnaissance de motif, ce composant DC n'est pas souhaitable car il donne une caractéristique qui change avec la valeur moyenne. Le choix d'une méthode multi-résolution comme le Log-Gabor peut être justifié par les propriétés des lignes de l'empreinte palmaire et par une base physiologique affirmant que, pour chaque emplacement d'une image palmaire, il existe des cellules couvrant au moins trois échelles sur un minimum de quatre octaves [1].

Dans la méthode proposée, un filtre log-Gabor 2D est appliqué aux images de ROI des quatre bandes spectrales (rouge, vert, bleu, NIR), permettant de générer 24 images avec différentes échelles et orientations (c'est-à-dire 4 échelles et 6 orientations). Ensuite, un codage compétitif binaire est appliqué pour conserver la valeur minimale pour chaque pixel de ces images, cette valeur minimale représente la valeur réelle d'intensité minimale des pixels de l'image filtrée (dans la même position spatiale). Pour construire l'image finale des caractéristiques, on a utilisé une représentation binaire basée sur le code Gray. Comme expliqué précédemment, l'image finale sera le résultat du codage compétitif appliqué sur les 24 images filtrées (4 orientations x 6 échelles). Chaque pixel de l'image finale sera remplacé par un équivalent binaire variant de 1 à 24 suivant son origine des 24 images filtrées. En général, le code Gray est utilisé pour éviter les états transitoires.

Pour la phase de décision nous avons utilisé une forme normalisée de l'indice de Jaccard, ce dernier est utilisé en statistiques pour comparer la similarité et la diversité entre des échantillons.



Résultats :

Il existe plusieurs bases de données proposant des images palmaires multi-spectrales. Parmi celles-ci, deux sont les plus connues : MS-Polyu et CASIA. MS-Polyu propose 6 000 images ROI sous quatre bandes différentes (rouge, vert, bleu, proche-infrarouge) provenant de 250 volontaires (195 hommes et 55 femmes). La base de données CASIA-MS-Palmprint contient 7200 images provenant de 100 volontaires. Pour évaluer les performances de notre méthode, nous avons utilisé la base de données MS-PolyU. Ce choix peut être justifié par la qualité des images et le nombre d'échantillons proposés. En effet, la base de données MS-Polyu propose directement des images ROI, comparée à CASIA qui propose une image entière de la main, ce qui nécessite l'implémentation d'une méthode de prétraitement pour extraire le ROI. MS-Polyu offre une image ROI avec une taille de 128 x 128, capturée avec un dispositif à couplage de charge (CCD). Ce dernier est composé d'une caméra, d'un convertisseur A/N, d'un objectif et d'une source de lumière multi-spectrale. Le convertisseur A/N est utilisé pour connecter l'ordinateur au CCD, la lumière multi-spectrale permet de contrôler le type d'image de sortie.

MS-PolyU offre deux bases d'images collectées dans deux sessions distinctes. chaque ensemble (3000 images) contient une sélection de 6 images de chaque main. Afin de bien organiser la base de données, nous avons considéré qu'une des collections représente la base de référence et la seconde et celle du test. Chaque image de caractéristiques générée depuis l'ensemble des références a été comparée avec l'ensemble des images des caractéristiques issues de la collection de test en utilisant la distance de Jaccard.

Nous avons utilisé comme critères d'évaluation, le ACC (précision de classification) en mode indetification et le EER (taux d'égale erreur) en mode vérification. Les tableaux ci-dessus illustrent les résultats obtenus par notre méthode comparée avec d'autres méthodes de l'état de l'art de type codage compétitif. En analysant ces résultats, on constate que la méthode proposée dépasse les méthodes de l'état de l'art pour les deux modes : vérification [2,3] et identification [4,5].

Méthode	Identification (ACC)
Zhang et al. [4]	98.4%
Raghavendra et al. [5]	95%
Méthode proposée	99.5%

Méthode	Vérification (EER)
Xu et al.[2]	0.0189%
Kisku et al. [3]	0.0248%
Méthode proposée	0.031%

Conclusion et perspectives :

Notre travail consistait à implémenter une nouvelle méthode de reconnaissance d'empreinte palmaire multi-spectrale en se basant sur le filtre Log-Gabor multi-échelles et le codage compétitif. Les résultats expérimentaux appliqués sur la base de données MS-PolyU ont montré que l'approche proposée permet d'améliorer les performances de reconnaissance dans les deux modes, identification et vérification. La reconnaissance biométrique basée sur des méthodes multi-échelles représente une réelle opportunité d'amélioration par rapport à ce qu'elle offre en terme de diversité des caractéristiques biométriques extraites, ce qui nous a amené à le choisir pour nos prochains travaux.

Références (5 max.):

- [1] R. L. De Valois, D. G. Albrecht, L. G. Thorell, "[Spatial frequency selectivity of cells in macaque visual cortex](#)", Vis. Res. Vol. 22, no. 5, pp. 545–559, 1982.
- [2] Y. Xu, L. Fei, J. Wen, D. Zhang, " [Discriminative and Robust Competitive Code for Palmprint Recognition](#) ", IEEE Transactions on Systems, Man, and Cybernetics, Vol. 48(2) ,2018
- [3] D. R. Kisku, A. Rattani, P. Gupta, J. K. Sing, C. J. Hwang, "[Human identity verification using multispectral palmprint fusion](#)", Journal of Signal and Information Processing, vol. 3, pp. 263--273, 2012.
- [4] S. Zhang, X. Gu, "[Palmprint recognition based on the representation in the feature space](#)", Optik-International Journal for Light and Electron Optics, vol. 124, no. 22, pp. 5434--5439, 2013.
- [5] N. Raghavendra, C. B. Norwegian, "[Novel image fusion scheme based on dependency measure for robust multispectral palmprint recognition](#)", Pattern Recognition, vol. 47, no. 6, pp. 2205--2221, 2014.
- [6] J. G. Daugman, "[Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters](#)". Journal of the Optical Society of America, pp. 1160-9, 1985.

Titre :

Iris Recognition via the use of Binarized Statistical Image Features and Hilbert Transform

Auteurs :

Nabil Hezil and Abdelhani Boukrouche

Résumé (100-200 mots) :

Binarized Statistical Image Features (BSIF) is a new and promising textural descriptor for human iris recognition. There are, however, some drawbacks when using only one descriptor, such as noise introduction and feature loss. In this paper, we argue that robust and secure biometric based iris recognition requires heterogeneous feature descriptors, and have therefore combined two successful image representations: BSIF and the Hilbert Transform. The descriptors are complementary in the sense that BSIF extracts information about the specific pattern to which pixels belong (stroke edge, stroke comers, stroke ends, inside the stroke or background) and the pattern orientation, while the Hilbert Transform encodes iris texture over a broader range of scales. In comparison with state of the art methods, our results indicate that the proposed method based on BSIF and the Hilbert Transform outperforms all the approaches considered in this study. The verification and identification performances of the proposed scheme were validated on three iris databases, namely the CASIA Iris Interval, IITD v1, and UBIRIS v1, in addition to a fused dataset. The use of the fused dataset showed the robustness of our method with respect to different iris pattern variations in terms of lighting, noise, multisensor and occlusion.

Mots-clefs (3 à 5 max.) :

Biometrics; Iris recognition; Image feature extraction; Automatic segmentation.

Contexte et état de l'art :

An iris recognition process consists of three major steps [1]. The first step is image acquisition, in which an image of the iris is captured. The second step is preprocessing: the main objective here is to remove any non useful information from the area where the iris is located, segmented, and normalized. The second step after image acquisition is image preprocessing. This is an important stage in the iris recognition system, as its goal is to remove the unwanted textures in the iris region such as those belonging to the pupil, eyelashes, eyelids and, light reflections in order to prepare the iris image for feature extraction with minimal noise. The preprocessing can be divided into two stages: iris localization and iris normalization.

Then, features are extracted to obtain the iris code template used for matching. We present in this work, an investigation on the use of the BSIF [2] descriptor to encode the iris texture. Given an image patch X of size $l \times l$ pixels and a linear filter W_i of the same size, the filter response s_i is obtained by:

$$s_i = \sum_{u,v} W_i(u,v)X(u,v) = w_i^T x \quad (1)$$

where the vectors w and x contain the pixels of W_i and X . The binarized feature b_i is obtained by setting $b_i = 1$ if $s_i > 0$ and $b_i = 0$ otherwise. To improve codification of the iris patterns, we use the Hilbert Transform to extract the phase information. It is well known that phase contains the structural information, which is an isotropic extension of the analytic image. This combination is advantageous because it can represent an iris image in terms of the local pattern, local orientation (BSIF) [3], local phase and local energy (Hilbert transform). The BSIF image contains only the real values where Hilbert Transform is used to define the complex valued analytic image associated with a real valued image (BSIF). Different settings and scenarios using several parameters and databases were tested and compared. A comparative evaluation with other state of the art methods using different iris databases will also be presented in table 1

Travail proposé (Décrire l'objectif du travail et indiquer clairement le problème/défi technique étudié) :

To the best of our knowledge, this is the first study to apply BSIF and the Hilbert Transform to extract texture features from an iris image. In order to provide an accurate recognition of individuals, the most discriminative information presented in an iris pattern must be extracted. The normalized iris image is divided into several regions to better characterize the details and to avoid the loss of spatial information. After performing several experiments with different values, we empirically observed that division of the iris image into 4 vertical non overlapping blocks and 3 horizontal non overlapping blocks is the best choice for this purpose. Then, the BSIF descriptor was applied to each block in the normalized iris image. Each pixel of the BSIF sub image contains data about the specific pattern to which the pixels belong (stroke edge, stroke comers, stroke ends, inside the stroke or background) and the pattern orientation. We tested different bases and we extracted the features at patch sizes of 3×3, 5×5, 7×7, 9×9, 11×11, 13×13, 15×15 and 17×17, Next, to extract the phase information which the iris contains, we consider the matrix T whose columns are computed by applying the Hilbert Transform in the obtained BSIF sub image and we compute the analytic signal (AS) defined as:

$$AS = BSIF \text{ sub image} + T$$

Finally, we generate the binary iris code IC (see Fig.3), as the logical index of all these components, within AS that has a positive imaginary part: Binary iris code (IC) = phase (AS) > 0

Résultats :

we evaluated the performance of the verification mode. We measured the performance in terms of Genuine Accept Rate (GAR), False Accept Rate (FAR), and False Rejection Rate (FRR) with the assumption that a test sample comes from a specific subject. For the CASIA Iris Interval database, the total number of comparison is 1542356, where the total number of intra class comparison is 10956, and that of interclass comparison are 1531400. For the IIT Delhi database, the total number of comparison is 1077828, where the total number of intra class comparison is 2016, and that of interclass comparison are 1075812. For the UBIRIS database, the total number of comparison is 116889, where the total number of intra class comparison is 2169, and that of interclass comparison are 114720. Fig.9 exhibits the receiver operating characteristic (ROC) curves of the proposed algorithm applied to the three databases in addition to the fused dataset. It can be observed that the proposed scheme achieves GAR of 91.02% at FAR=10-2%, whereas the Li code [4] achieves GAR of 89.93%, and the Masek code [5] gives GAR of 88.82% on the fused dataset, that contains iris images with irregularities due to motion blur, off angle gaze deviation, diffusion, and other real world problems.

The distance between the minimum Hamming distance value for inter-class comparisons and maximum Hamming distance value for intra-class comparisons can be used as a metric to measure the separability potency see fig 5. However, this is not a very accurate measure since outliers will corrupt the value calculated, and the measure is dependent on the number of iris templates compared. Thus, the better metric is decidability (d'), which takes into account the mean and standard deviation of the intra-class and inter-class distributions.

We conducted a one to many search within the entire database for a given test sample to find the template that best matches with a given test sample see fig 4.

Conclusion et perspectives :

In this work, we have presented an approach for automatic iris recognition based on Binarized Statistical Image Features Hilbert (BSIFH), which is an isotropic extension of the analytic signal. Three different databases were analyzed for this particular study: the popular CASIA Iris Interval, UBIRIS v1, and IITD databases, and the results were compared with some well known feature extraction approaches under the same conditions. The proposed approach demonstrated its ability to attain better results compared to well known methods reported in the literature. To further analyze the system, a combined database was created from the above three databases. It was found from experimental results that the proposed iris recognition algorithm outperformed the state-of-the-art methods in terms of recognition rate. We can note that the BSIF filters are learned from random textured images. Consequently, one can expect better results by learning new BSIF filters from iris images. It is also of interest to apply adaptive boosting (AdaBoost) or deep learning algorithms to automatically find the best features and parameters instead of empirical experiments selection. as well to fuse our iris recognition method with other modalities to design a multimodal biometric system that overcomes the weaknesses of the unimodal system.

Références (5 max.):

- [1] Daugman, J.: 'How iris recognition works', IEEE Trans. Circuits Syst. Video Technol. (2004),14(1) 21–30
- [2] Kannala J., Rahtu E.: 'BSIF: binarized statistical image features', In Proc. IEEE Int. Conf. on Pattern Recognition. (2012). pp. 1363–1366.
- [3] Hezil, N. Boukrouche, A.: 'Multimodal biometric recognition using human ear and palmprint', IET Biometrics. (2017). 6(5), pp. 2047-4946.
- [4] Li X.: 'Modeling intra-class variation for non-ideal iris recognition', In Int. Conf. On Biometrics. (Springer Berlin Heidelberg) (2006). 419–427.
- [5] Libor M.: 'MATLAB Source Code for a Biometric Identification System Based on Iris Patterns', The School of Computer Science and Software Engineering, the University of Western Australia (2003).

Il est recommandé d'insérer des images/figures et des tableaux afin d'illustrer le propos.

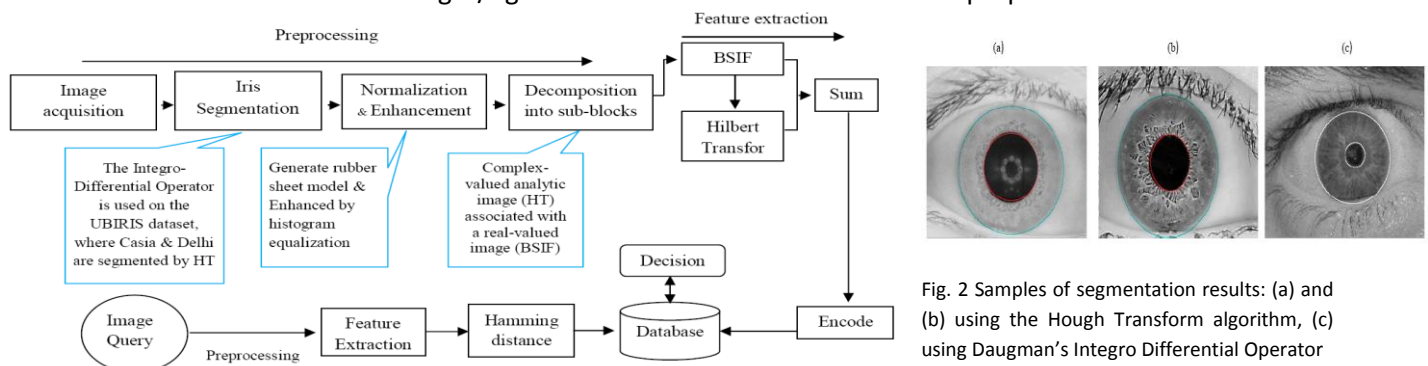


fig. 1 Schematic diagram of the proposed iris biometric recognition method.

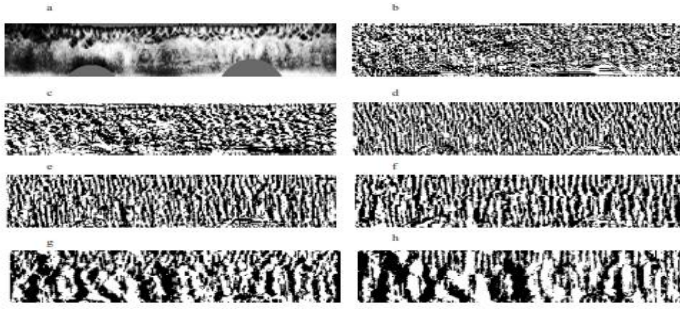


Fig. 3 Iris code with various patch sizes: (a) Normalized iris (b) - (h) corresponding iris code with BSIFH patch sizes of 5×5 , 7×7 , 9×9 , 11×11 , 13×13 , 15×15 , 17×17 and a basis length of 12.

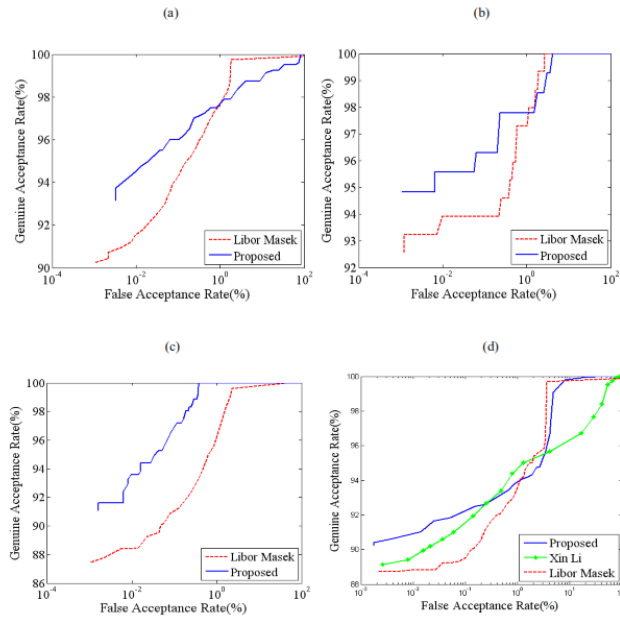


Fig. 4 ROC curves showing a comparison of different existing recognition techniques on (a) CASIA, (b) IITD, (c) UBIRIS, and (d) the fused dataset.

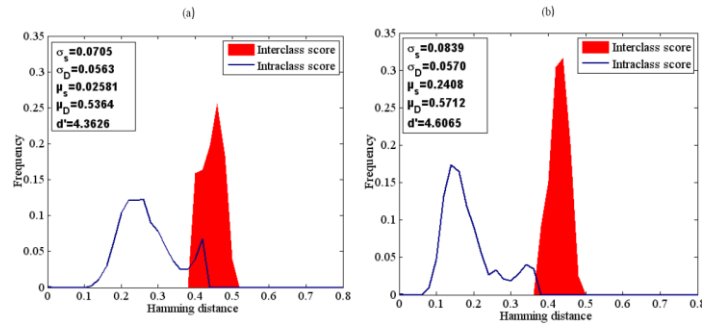


Fig. 5 The experimental imposter and genuine similarity score distributions obtained during the recognition tests with (a) Masek's algorithm and (b) our proposed algorithm on the fused dataset.

Algorithm	Correct Recognition Rate (%)			
	CASIA	IITD	UBIRIS	fused
Libor	97.52	97.18	97.03	96.41
Li	/	/	/	95.77
Roy et al.	97.21	/	/	/
Liu et al.	97.11	/	/	/
Umer et al	/	97.40	97.51	/
Tsai et al.	/	/	97.20	/
Sundaram et al.	/	/	97.00	/
Erbilek et al.	/	95.83	/	/
Ma et al.	94.90	/	/	/
Dhage et al.	/	88.56	/	/
Proposed*	98.86	98.33	97.45	97.34

Table 1: CRRs Comparisons (our algorithm vs. some state of the art approaches)



Titre :

Performance of no-reference image quality metric for face images-based biometric systems

Auteurs :

Xinwei Liu^{1,2}, Christophe Charrier¹, Marius Pedersen² and Patrick Bours²

¹Normandie Univ., UNICAEN, ENSICAEN, CNRS, GREYC, Caen, France

²NTNU - Norwegian University of Science and Technology, Gjøvik, Norway

Résumé (100-200 mots) :

The accuracy of face recognition systems is significantly affected by the quality of face sample images. There are many existing no-reference image quality metrics (IQMs) that are able to assess natural image quality by taking into account similar image-based quality attributes. Previous study showed that IQMs can assess face sample quality according to the biometric system performance. In addition, re-training an IQM can improve its performance for face biometric images. However, only one database was used in the previous study, and it contains only image-based distortions. In this paper, we propose to extend the previous study by use multiple face database including FERET color face database and apply multiple setups for the re-training process in order to investigate how the re-training process affect the performance of no-reference image quality metric for face biometric images. The experimental results show that the performance of the appropriate IQM can be improved for multiple databases, and different re-training setups can influence the IQM's performance.

Mots-clefs (3 à 5 max.) :

Biometric quality assessment, Face recognition, Image quality metric, Machine Learning ,performance evaluation

Contexte et état de l'art :

The face become one of the most common and successful modalities for biometric recognition in the past decade. However, face recognition is still a challenging issue when degraded face images are acquired. Low sample quality is a main reason for matching errors in biometric systems and may be the main weakness of some applications. But the number of quality assessment methods dedicate for face images is limited. Meanwhile, there are many existing IQMs that have been developed for natural images' quality. It is interesting to investigate whether such IQMs can be used for the assessment of face biometric images. There are very few researches investigate such issue and previous studies evaluated no-reference IQMs by using only one specific face database. The existing works show that the performance of IQMs can be improved by using high quality face images for re-training. However, there is only one re-training setup. Therefore, it is necessary to conduct new experiments to investigate this interesting topic.

Travail proposé (Décrire l'objectif du travail et indiquer clairement le problème/défi technique étudié) :

In order to investigate how re-training process affect the performance of no-reference image quality metric for face images, we propose to 1) use well selected 15 no-reference IQMs for the evaluation process, 2) use two different databases to conduct the experiments, 3) use face images, iris images, and face+iris images for the re-training setups.

The proposed approach is briefly describe hereafter:

1. We use two face databases: the first one is GC₂-multi-modality biometric image quality database. We only use face images from it. Three cameras are used for the acquisition: 1) a Light Field Camera (LFC), 2) a smartphone camera, and 3) a reflex camera. 50 subjects participated in the acquisition. 2250 (15*50*3) raw face images are obtained in the database. In addition, different types of degradations are introduced in the database. Each face image is degraded into five distortion levels for eight degradations. There are totally 92250 face images in GC₂ database. The second face database is FERET color face database. It has 269 subjects and there are two acquisition sessions for most of subjects. Totally 14126 face images are in this database.
2. 15 no-reference IQMs are selected for the performance evaluation (see Table hereafter). In this study, we select the best performed IQM to conduct the re-training process on above mentioned three setups for both GC₂ and FERET databases.

IQMs	Distortion-specific	Generalized purposes
Natural Scene Statistics (NSS)-based	CONTRAST	BIQI, BLIINDS2, BRISQUE, ILNIE2
Non NSS-based	JNBM, DCTSP, SH, CONTRAST2, JPEG, PWN	AQI, AQIP, diplQ, SSEQ

3. The face recognition system used in this paper is 'The PhD (Pretty helpful Development functions for) face recognition toolbox', which is a collection of Matlab functions and scripts for face recognition.
4. Rank-ordered detection error trade-off (DET) curve is one of the most commonly used method to evaluate the performance of quality assessment approaches. It plots false none match rate (FNMR) versus false match rate (FMR). Quality-bin-based approaches also used in this work. If a certain percentage of low quality samples are excluded from the dataset, the biometric system performance would become 'better' (e.g. Equal Error Rate (EER) (when FMR and FNMR are equal) would decrease). We use DET curve and EER as two methods to represent the performance of IQMs. We omit the percentile low quality samples and keep 80%, 60%, and 40% of highest quality samples from each subject from the quality scores given by IQMs.

Résultats :

1. The overview of the performance of 15 selected IQMs is given in Table 2. We can see that only ILNIQE2 can improve the

system performance for all cameras. So we will conduct the following experiments only on ILNIQE2. For the moment, the re-training is only using face and face+iris images. The results from using iris images for re-training will be added to the final paper.

- We use EER values by omitting lowest quality face sample one by one until only one highest quality face sample is left from each subject as an indicator for GC₂ database. The comparison of the change of EER values between two setups is illustrated in Figure 2. The performance of re-trained ILNIQE2 by using face+iris images is better before 60% of high quality samples are still kept in the database. However, after half of the lowest quality face images are omitted, the EER becomes 0 for the blue line. The red line has an increasing at the end. It means that the overall performance of re-trained ILNIQE2 using face+iris is not as good as when only use face images for re-training.
- We now use FERET color database to evaluate the performance of ILNIQE2. The training datasets are face images from GC₂ database. We will present the results of DET curves in the final presentation. Here, we only illustrate EER values by omitting lowest quality face sample one by one in Figure 3. The re-trained ILNIQE2 using only face images has the overall better performance. The performance of re-trained ILNIQE2 using face+iris images has better performance than the original one, but becomes worse when only 40% of high quality samples are left in the database. In addition, EER values from different setups are very similar between 70% and 45% of high quality samples are left in the database.
- In order to discover why the performance of re-trained ILNIQE2 decreased when adding iris images into the training dataset, we will use ILNIQE2 to assess the quality of face, iris, and face+iris images from the training dataset in the final paper. The distribution of the quality scores can be used for the analysis of why the performance changed.

Table 2. Overview of the performance of selected IQMs for face images. The green cells in the table represent that by using such metrics to omit lowest quality face sample one by one until only one highest quality face sample left from each subject, the system performance is improved. The yellow cells in the table represent the metrics are neutral, which have little influence to the system performance. The red cells in the table represent the metrics decrease the performance of face recognition system.

	LFC	Phone	Reflex
AQI			
AQIP			
BIQI			
BLINDS2			
BRISQUE			
CONTRAST			
DCTSP			
ILNIQE2			
JNBM			
PWN			
SH			
SSEQ			
CONTRAST2			
dipIQ			
JPEG			

Improved performance
Neutral
Decreased performance

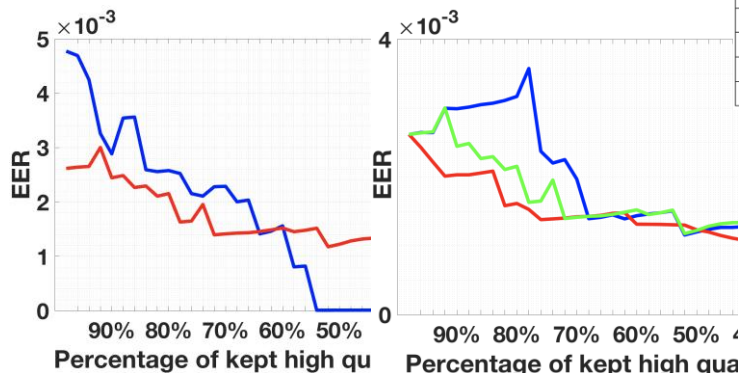


Figure 2. Comparison of EER by omitting lowest quality sample one by one using re-trained ILNIQE2 for each subject between the two setups. The x-axis represents the percentage of kept high quality face samples. The blue line represents the re-trained ILNIQE2 by using only face images, and the red line represents the re-trained ILNIQE2 using both face and iris images. The y-axis represents the EER value.

Figure 3. Comparison of EER by omitting lowest quality sample one by one using re-trained ILNIQE2 for each subject between the three setups.

Conclusion et perspectives :

The novelty of this work is that we setup experiments to investigate how re-training process affect the performance of no-reference IQMs for face images. Compared to the existing works in the field, this paper 1) select more IQMs (15) for the evaluation, 2) use two databases for the evaluation, 3) introduce distortions to face images, 4) use three (face, iris, face+iris) different re-training setups to investigate the performance of IQMs, 5) analyze why different re-training setups could affect IQMs' performance on face images. The results of this paper can be further used as a reference for the optimization of IQMs on biometric data.

Références (5 max.):

- [1] Jain, Anil K., and Stan Z. Li. Handbook of face recognition. New York: Springer, 2011.
- [2] Liu, Xinwei, Marius Pedersen, Christophe Charrier, and Patrick Bours. "Can No-Reference Image Quality Metrics Assess Visible Wavelength Iris Sample Quality?." In IEEE International Conference on Image Processing. 2017.
- [3] Rana, Muhammad Ehsan, Ahmad Afzal Zadeh, and Ahmad MohammadMahmood Alqurneh. "Use Of Image Enhancement



Titre :

Joint optimization of lifting operators for stereo image coding

Auteurs :

Ismail Bezzine⁽¹⁾, Mounir Kaaniche⁽¹⁾, Noor Al-Maadeed⁽²⁾, Saadi Boudjit⁽¹⁾, Azeddine Beghdadi⁽¹⁾

⁽¹⁾L2TI, Institut Galilée, Université Paris 13, Sorbonne Paris Cité, 99 Avenue Jean Baptiste Clément, 93430 Villetaneuse, France

⁽²⁾Qatar University, Ibn Khaldoun Hall, Doha, Qatar

Résumé (100-200 mots) :

Recent advances in stereoscopic systems have extended the range of our experience and contributed to a widespread usage of stereo images. The great interest in these systems has resulted in huge amount of data which needs to be compressed for storage and transmission purposes. In this context, vector lifting scheme has been found to be an efficient approach for stereo image coding. Therefore, the coding performance depends on the design of the involved lifting operators referred to as prediction and update filters. This paper proposed to build content-adaptive decomposition based on previous Non-separable vector lifting scheme (NS-VLS) decomposition through joint optimization technique. While a weighted L-1 minimization technique has already been investigated for stereo image coding and for hologram compression, this work aims at developing and studying a new joint optimization strategy for the design of all the involved filters used with the left and right images. While sparse optimization algorithms have been recently employed for still image coding, this paper deals with their extension to the context of stereo image coding. An independent L-1 and weighted L-1 based optimization techniques as well as a joint optimization of both images lifting operators are investigated.

Mots-clefs (3 à 5 max.) :

Stereo images, adaptive coding, vector lifting scheme, non-separable transform, optimization.

Contexte et état de l'art :

Stereoscopy is one of the growing research areas in computer vision and image processing. In fact, a stereoscopic system consists in generating two images by recording the same scene from two slightly different view positions. The obtained images, referred to as left and right images, are then merged by the brain to perceive the scene in three dimensions. Thanks to its simplicity, the stereovision has been widely used in various application fields. Therefore, due to the increasing interest in stereo images, it becomes mandatory to handle the huge amounts of data generated from various stereoscopic systems. Thus, it is compulsory for practical use to design efficient stereo image coding schemes with high visual information transmission quality and low storage capacity.

A basic approach for stereo images coding may consist in encoding independently the left and right views by employing existing still image encoders. However, such approach may not appear so efficient since it does not exploit the main characteristics of these images. In fact, stereo images present similar contents and exhibit a high correlation since they correspond to the same 3D scene. Therefore, efficient stereo image coding schemes could be designed by exploiting the inter-view redundancies. In literature, most of the proposed stereo images coding methods are based on the conventional scheme. The latter consists on encoding the reference and residual images as well as the disparity information [1].

In this paper, we were especially interested on an approach that consists in using a multiscale decomposition based on the concept of vector lifting scheme (VLS). Note that, unlike conventional lifting scheme, the VLS is a joint wavelet decomposition that aims at exploiting the inter-view correlations to generate two compact multi-resolution representations of the left and right images. While a separable decomposition has been carried out in [2], its extended non separable version (NS-VLS) has been developed in [3]. The design of the lifting operators of both views for a NS-VLS decomposition is based on sparse optimization algorithms using different ℓ_1 based minimization techniques. It is important to note here that different sparse optimization algorithms have been employed for still image coding [4] while only a weighted ℓ_1 minimization technique has already been investigated in [3] for stereo image coding and in [5]

hologram compression. In fact, if we focus on the lifting structure applied to the left image, it can be seen in Fig. 1 that the diagonal detail coefficients resulting from the first prediction step are used in the second and third prediction steps to generate the left detail coefficients. Therefore, instead of minimizing the ℓ_1 -norm of the diagonal detail coefficients, it becomes more interesting to optimize the first prediction filter by minimizing a weighted sum of the ℓ_1 -norms of the three detail subbands of the left image [4]. However, this method is applied on each view separately and the high inter-view correlations between the two views are not taken into consideration for the design of the filters.

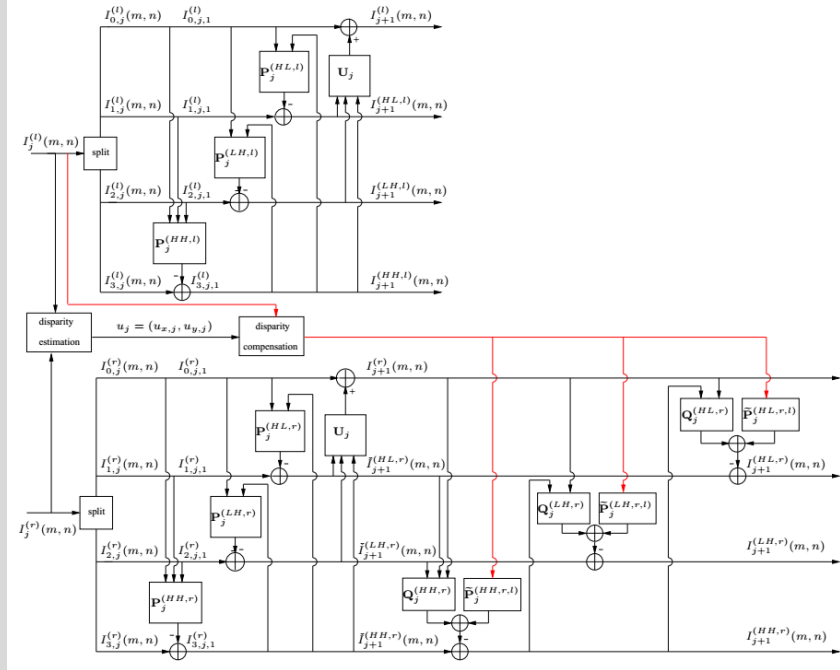


Figure 1 : NS-VLS decomposition structure

Travail proposé (Décrire l'objectif du travail et indiquer clairement le problème/défi technique étudié) :

According to the NS-VLS decomposition, the lifting stage applied to the left image is similar to the first stage applied to the right image used to generate the approximation subband and three intermediate detail subbands. Moreover, one of the main characteristics of stereo images is that they present high inter-view correlations since they correspond to the same 3D scene. Therefore, instead of optimizing each view independently of the other one, it becomes interesting to design a joint optimization approach to take into account the previous observations.

To exploit the correlation existing between the left and right, we propose first to assume that the filters of the first lifting stage employed with the right image similar to those used with the left image referred as the intra prediction filters. Thus, for notations concision, the three intra-prediction filters as well as the update one will be simply denoted by $P_j^{(o)}$ and U_j .

$$\begin{cases} P_j^{(HH,r)} = P_j^{(HH,l)} = P_j^{(HH)} \\ P_j^{(HL,r)} = P_j^{(HL,l)} = P_j^{(HL)} \\ P_j^{(LH,r)} = P_j^{(LH,l)} = P_j^{(LH)} \\ U_j^{(r)} = U_j^{(l)} = U_j \end{cases}$$

Moreover, since these filters are applied both to the left and right images, we also propose to design a hybrid weighted ℓ_1 criterion defined simultaneously on the stereo pairs. More precisely, this criterion is the weighted sum of the ℓ_1 -norm of the three detail subbands of the left image as well as the three intermediate detail subbands of the right one. Therefore, the new hybrid weighted ℓ_1 criterion can be expressed as follows:

$$\forall o \in \{HL, LH, HH\}$$

$$\mathcal{J}_{w\ell_1}^{(r,l)}(P_j^{(o)}) = \sum_{o \in \{HL, LH, HH\}} \sum_{m=1}^{M_j} \sum_{n=1}^{N_j} \left(\frac{1}{\alpha_{j+1}^{(o,l)}} |I_{j+1}^{(o,l)}(m, n)| + \frac{1}{\alpha_{j+1}^{(o,r)}} |I_{j+1}^{(o,r)}(m, n)| \right)$$

Similarly to the weighted ℓ_1 criterion proposed in [3], the new hybrid one will also be minimized using the Douglas-Rachford algorithm in a product space. Once the prediction filters used in the first lifting stage of both views have been jointly optimized, the last three inter prediction filters used with the right image are optimized by minimizing the weighted sum of the ℓ_1 -norms of the three detail subbands of the right image.

Résultats :

To show the performance of the proposed optimization method, we will consider the following ones carried out over three resolution levels: The first one called “independent” consists in coding independently the left and right images by applying the 5/3 transform to each view, the second one called “NS-VLS-OPT-L1” corresponds to the NS-VLS where the prediction filters are optimized separately by minimizing the ℓ_1 -norm of the detail coefficients, the third one designated by “Residual” consists in coding the left image and the residual one by using the 5/3 wavelet transform. and the last one designated by “NS-VLS-OPT-WL1-Joint” is the proposed joint optimization method.

All these approaches are compared in terms of rate-distortion performance. Fig. 2 illustrates the variations of the PSNR versus the bitrate for the “houseof” stereo images. We can easily observe that the gain is about 0.1-0.2 dB compared to the standard ℓ_1 minimization technique and up to 0.4 dB compared to the independent approach. Moreover, compared to the residual method, the joint optimization algorithm leads to a gain of about 0.3-0.7 dB in terms of PSNR for high bitrates.

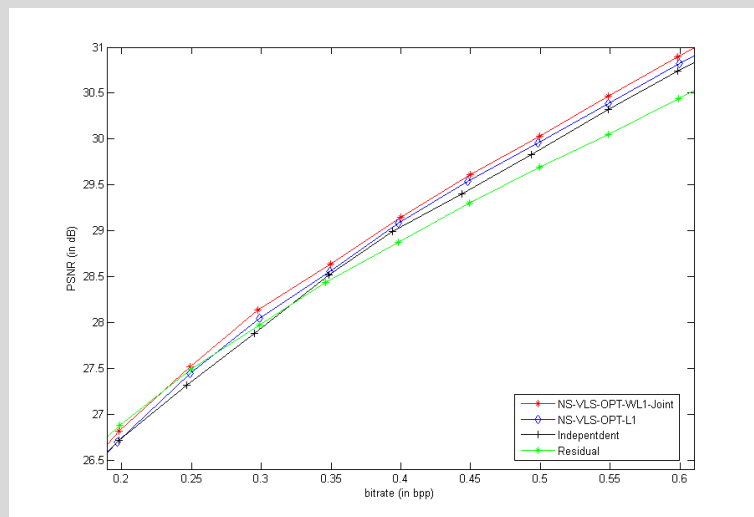


Figure 2: PSNR (in dB) versus the bitrate (bpp) after JPEG2000 encoding for the “houseof” stereo pair.

Conclusion et perspectives :

In this paper, we proposed a joint optimization technique to design the prediction filters of a non-separable vector lifting scheme for stereo image coding purpose. The experimental results carried out on different real stereo images have shown the efficiency of this method. In a future work, we plan to extend this work to the context of multiview coding.

Références (5 max.):

- [1] M. S. Moellenhoff, M. W. Maier, Characteristics of disparity-compensated stereo image pair residuals, *Signal Processing: Image Communications* 14 (1998) 49-55.
- [2] M. Kaaniche, A. Benazza-Benyahia, B. Pesquet-Popescu, J.-C. Pesquet, Vector lifting schemes for stereo image coding, *IEEE Transactions on Image Processing* 18 (11) (2009) 2463-2475
- [3] M. Kaaniche, B. Pesquet-Popescu, J.-C. Pesquet, ℓ_1 -adapted non separable vector lifting schemes for stereo image, in: *European Signal Processing Conference*, Bucharest, Romania, 2012, p. 5 pages
- [4] M. Kaaniche, A. Benazza-Benyahia, B. Pesquet-Popescu, J.-C. Pesquet, Non separable lifting scheme with adaptive update step for still and stereo image coding, *Elsevier Signal Processing: Special issue on Advances in Multirate Filter Bank Structures and Multiscale Representations* 91 (12)
- [5] Y. Xing, M. Kaaniche, B. Pesquet-Popescu, F. Dufaux, Sparse based adaptive non separable vector lifting scheme for holograms compression, in: *International Conference on 3D Imaging*, Liège, Belgium, 2015, p. 8 pages.



Titre :

Reference list optimization for improvement of compression of 2D-multiview content for free-navigation applications.

Auteurs :

Pavel Nikitin, Marco Cagnazzo, Joël Jung

Résumé (100-200 mots) :

One of the most challenging goals of future immersive services is to enable the observation of a scene from any view-point, thus making free-navigation possible under certain constraints. In order to provide such kind of services with smooth navigation, a huge amount of views should be available at the client's device, in particular for the case of 2D-multiview content, where cameras are positioned on a 2D grid in order to provide both horizontal and vertical parallax. This kind of content requires of course a large coding rate, therefore improving the compression performance of video encoders is particularly relevant in this case. This paper studies how the reference lists and their parameters affect the compression, by taking into account the spatial position of each camera. Three parameters are addressed in this work: the number of reference lists, the number of reference pictures and the ordering of pictures in reference lists. An average of **11.0%** bitrate saving is achieved compared to the state of the art.

Mots-clefs (3 à 5 max.) :

Compression, free-navigation, 2D multiview, reference pictures lists

Contexte et état de l'art :

In this study we consider a non-normative modification of MV-HEVC [1] related to the reference list construction and to the number of reference pictures in reference lists. The optimization of the reference list for temporal prediction has been widely studied for "classical" (ie., single view) video content [2]; likewise, there has been some works on the selection of reference view for optimizing inter-view prediction in the case of horizontal-parallax multiview video: for example, Maugey et al. [3] have investigated in details the case of 1D-multiview content and the free navigation scenario along the 1D path. However, in the case of 2D-multiview (i.e. content with horizontal and vertical parallax), the optimization of the reference list structure has been overlooked in the literature. In one of the few works dealing with this topic [4], it is claimed that serpentine coding order among views gives the best RD performance, but no proof is given.

In this work we study the impact of the reference lists construction on the RD performance. In order to evaluate it, we set as anchor an MV-HEVC encoder with direct-P inter-view structure, one reference list, one reference image in the list and serpentine coding order, as shown in Figure 1. Our anchor fits to the MPEG-I's common test conditions [1]. In particular, the texture and depth frames are compressed separately, and after decoding texture and depth information, virtual views on the positions of captured views are synthesized in order to test how the compression impacts on the quality of the synthesis. In the results, we report the BD-rate difference compared to the anchor. The final results consist of PSNR of compressed texture over the total bitrate (video Bjontegaard Delta Rate, video BD rate for short) and PSNR of synthesized views over the total bitrate (synth BD-rate for short).

For this study we have used five 2D-multiview sequences captured by 4x4 cameras: Technicolor Painter, ULB Unicorn, Orange Shaman, Orange Dancing, and Orange Kitchen.

Travail proposé (Décrire l'objectif du travail et indiquer clairement le problème/défi technique étudié) :

Our goal is define how to achieve an effective reference list configuration, based on camera positions. Firstly we have tested the impact on BD-rate performance of increasing the number of reference lists from one to two, thus allowing the encoder to use the bi-prediction. We have obtained **0.5%** bitrate saving. Secondly we have tested the dependency of BD-rate on the number of reference pictures in reference lists. The results are depicted in Figure 3. Finally we have modified the way we build the reference lists. We calculate the distance

between the camera that corresponds to the current view and cameras that correspond to all already decoded views, and the references views are put the lists in increasing distance order. The tie-break rules are the following: if the equidistant views are horizontal and vertical neighbors of the current view (e.g. V6 and V9 references for V10, see Fig.1), the order is H-V in L0 and V-H in L1. Otherwise they are ordered putting the last coded view first (e.g., V5 and V7 for V10, see Fig.1). The proposed configuration is compared with some basic reference list organizations, where we have put the pictures in reference list either in ascending (Asc) or descending (Desc) coding order. As an example, the complete reference lists for the view 10 are depicted in Table 1 for the basic and the proposed methods.

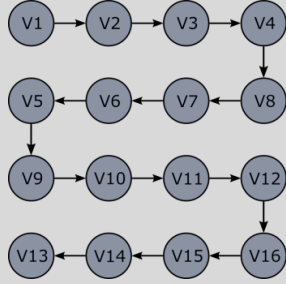


Figure 1. Serpentine coding order.

Configuration	Reference list	Reference pictures
Asc Asc	L0	V1 V2 V3 V4 V8 V7 V6 V5 V9
	L1	V1 V2 V3 V4 V8 V7 V6 V5 V9
Desc Desc	L0	V9 V5 V6 V7 V8 V4 V3 V2 V1
	L1	V9 V5 V6 V7 V8 V4 V3 V2 V1
Asc Desc	L0	V1 V2 V3 V4 V8 V7 V6 V5 V9
	L1	V9 V5 V6 V7 V8 V4 V3 V2 V1
Desc Asc	L0	V9 V5 V6 V7 V8 V4 V3 V2 V1
	L1	V1 V2 V3 V4 V8 V7 V6 V5 V9
Proposed	L0	V9 V6 V5 V7 V2 V8 V3 V1 V4
	L1	V6 V9 V5 V7 V2 V8 V3 V1 V4

Table 1. Examples of reference list ordering for V10.

Résultats :

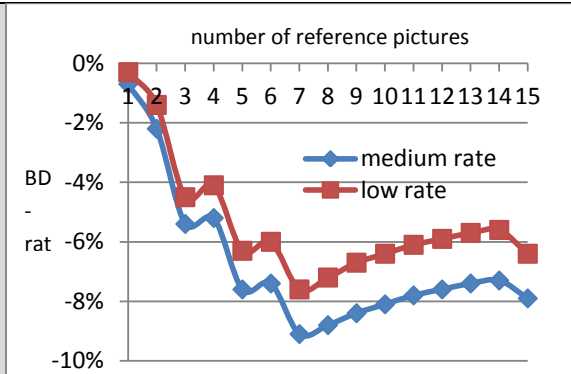


Figure 3. Dependency of BD-rate on number of references. Using 7 references provides the best RD performance

Configuration	video BD-rate		synth BD-rate	
	medium	low	medium	low
desc desc	-10.6%	-9.3%	-13.0%	-11.8%
asc asc	-9.1%	-7.6%	-9.2%	-9.2%
asc desc	-7.1%	-5.3%	-8.6%	-7.7%
desc asc	-10.3%	-9.1%	-11.7%	-10.1%
proposed	-11.0%	-9.7%	-12.6%	-12.6%

Table 2. Dependency of BD-rate on reference list ordering.

Conclusion et perspectives :

In this paper we have extensively studied the impact on the performance of compression based on three parameters of reference list organization. By using two reference lists we benefit from bi-prediction and can achieve **0.5%** BD-rate gain. By applying the proposed method of construction reference lists we have obtained **11.0%** BD-rate gain for medium bitrate and **9.7%** for low bitrate and **12.6%** for synthesized views. Thus, putting reference pictures in the list starting from the closest to the current view provides the best BD-rate gains. The usage of seven references provides the best performance for compression. Using lists with larger number of references increases the signaling cost without being compensated by better prediction. Future work will be devoted to study the impact of reference list characteristics in the case when both temporal and interview predictions are used.

Références (5 max.):

- [1] CTC on 3DoF+ and Windowed 6DoF, J. Jung, B. Kroon, R. Doré, G. Lafruit, J.Boyce. w17726, July 2018 Ljubljana, Slovenia.
- [2] Hierarchical B pictures, H. Schwarz, H. Schwarz, and T. Wiegand. Joint Video Team, Doc. JVT-P014, 2005.
- [3] Reference view selection in DIBR-based multiview coding, T. Maugey, G. Petrazzuoli, P. Frossard, M. Cagnazzo, and B. Pesquet-Popescu. IEEE Transactions on Image Processing, 21:1808-1819, 2016.
- [4] Subjective evaluation of super multi-view compressed content on high end lightfield 3D display, A. Dricot, J. Jung, M. Cagnazzo, B. Pesquet, F. Dufaux, P. Kovacs, and V. Kiran Adhikarla. Elsevier Signal Processing: Image Communication, 39:369-385, November 2015.



Titre :

Décomposition multi-échelles de maillages 3D hexaédriques dans le domaine des géosciences.
Étude des performances en compression sans pertes.

Auteurs :

BOUARD Lauriane, DUVAL Laurent, PAYAN Frédéric, ANTONINI Marc

Résumé (100-200 mots) :

Des méthodes de simulation employant des maillages sont mises en œuvre dans de nombreux champs scientifiques pour caractériser des phénomènes physiques sous-jacents. Les besoins croissants en précision induisent l'utilisation de maillages toujours plus volumineux, et entraînent des problèmes de visualisation, de manipulation ou de stockage des données. En géosciences, d'immenses zones géologiques sont modélisées par des maillages hexaédriques géométriquement complexes, émulant des propriétés physiques du sous-sol. En combinant différents types d'ondelettes classiques ou morphologiques qui préservent les discontinuités, HexaShrink (HS) apporte à cette problématique une solution multi-échelle cohérente. Dans cet article, nous analysons les performances de compression de HS, par l'étude exhaustive de sept maillages hétérogènes. Globalement, les taux de compression accrus, obtenus grâce à la décomposition, sont satisfaisants. Néanmoins, l'analyse distincte des différentes composantes définissant les maillages (géométrie, propriétés) révèle que l'utilisation d'encodeurs génériques n'est pas toujours optimale, ouvrant des perspectives vers des encodeurs multi-échelles plus à même d'exploiter les structures intrinsèques des niveaux de détail.

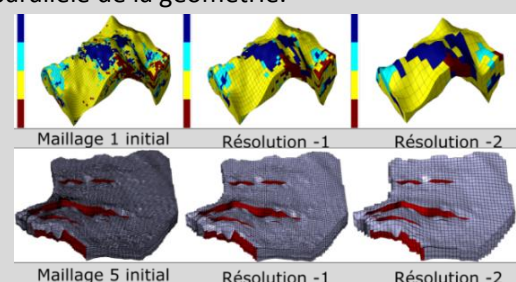
Mots-clefs (3 à 5 max.) :

compression, décomposition multi-échelle, maillages volumiques, ondelettes, simulation

Contexte et état de l'art :

Les simulations scientifiques utilisent des maillages de tailles croissantes [1], compliquant leur visualisation, leur transfert, ainsi que leur stockage. HS [2] propose une solution de décomposition multi-échelle pour les maillages hexaédriques utilisés en géosciences (*Corner Point Grid*), qui intègrent des discontinuités (cf. figure, *maillage 5 faillé*, *surfaces rouges*). De plus, de nombreuses propriétés physiques sont généralement associées à chaque cellule hexaédrique, composantes qu'il faut traiter en parallèle de la géométrie.

Pour chacune une ondelette adaptée génère une structure à différentes résolutions incluant approximations et détails, tout en préservant l'aspect visuel des discontinuités (cf. figure). Ce processus complètement réversible [3] traite de façon cohérente la géométrie et les différents types de propriétés : activité, variables physiques continues et catégorielles. Des travaux antérieurs ont démontré un potentiel initial de HS pour la compression sans perte de ces maillages [3].



Travail proposé (Décrire l'objectif du travail et indiquer clairement le problème/défi technique étudié) :

Nous présentons ici une étude exhaustive des performances en compression de HS associé à différents encodeurs génériques (gzip, bzip2, LZMA), dont l'objectif est d'en optimiser l'efficacité. Les maillages pouvant avoir des propriétés diverses, nous étudions un panel représentatif de 7 maillages hétérogènes aux dimensions et complexités variables. Dans un premier temps, le calcul des gains de compression pour chaque résolution de maillage évalue les performances globales de HS associé à un encodage sans perte. Dans un second temps les résultats sont détaillés pour chaque composante du maillage, afin d'évaluer l'efficacité relative des méthodes de chacune. Dans un format *Corner Point Grid* on considère 5 types d'objets. Les piliers (PILLAR), associés à des coordonnées de profondeur (ZCORN), constituent la géométrie du maillage. Des attributs associés aux cellules définissent les propriétés physiques du maillage — continues (CONTI) ou catégorielles (CATEG) — ainsi qu'une propriété booléenne d'activation des cellules (ACTNUM).

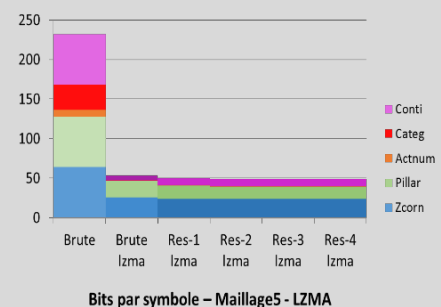
Résultats :

#	Caractéristiques			Taux de compression					
	Hexaèdres	Failles	Taille du fichier	gzip		bzip2		LZMA	
				Direct	HS	Direct	HS	Direct	HS
1	493 600	Non	4,62 MB	3,73	5,67	4,98	6,13	6,43	7,44
2	1 000 000	Non	42,46 MB	3,23	7,58	8,41	13,03	10,12	13,35
3	36 816	Oui	1,46 MB	2,67	4,05	2,99	4,93	3,63	5,48
4	210 000	Oui	7,88 MB	1,83	2,76	1,89	3,23	2,21	3,65
5	450 576	Oui	22,73 MB	2,46	3,26	2,55	2,92	3,33	3,81
6	5 577 325	Oui	274,57 MB	2,31	4,24	2,25	4,68	3,04	5,73
7	13 947 600	Oui	580,94 MB	3,20	6,72	5,98	10,12	12,52	10,23

Le taux de compression global d'un maillage correspond au rapport de sa taille initiale divisée par la taille de ce même maillage encodé directement ou après décomposition par HS. On constate un gain dû à la décomposition en ondelettes par HS quel que soit l'encodeur utilisé (*voir tableau*). Prenons l'exemple du maillage 5 pour illustrer nos propos, avec un encodeur gzip : le taux de compression augmente de 2,46 à 3,26. Le gain est d'autant plus grand que l'encodeur est récent : le taux de compression avec un encodeur LZMA passe de 3,33 pour une donnée brute à 3,81 pour une donnée décomposée.

Au-delà de la performance globale, et afin d'améliorer notre méthode, on s'intéresse maintenant aux détails des performances pour chaque composante du maillage. L'histogramme ci-dessous présente pour chacune d'elles le nombre de bits moyen nécessaire pour coder une de ses valeurs. La donnée géométrique représente 3/4 de la taille du maillage, la compression de cette composante a donc un énorme impact sur le résultat final. Par exemple, une coordonnée de profondeur (ZCORN) est initialement codée sur 65 bits.

LZMA réduit ce nombre à 19 bits, et descend jusqu'à 16 bits si on utilise HS avec deux niveaux de décomposition. Au-delà de deux niveaux de décomposition les tests montrent que les gains de compression deviennent négligeables. Cependant malgré de bons taux de compression et l'aspect multi-échelle d'HS, les encodeurs génériques testés n'exploitent pas les redondances inter-échelles. Ainsi pour une propriété continue (CONTI), LZMA ne tire pas bénéfice de la décomposition d'HS : leur association augmente le nombre de bits par valeur (comparé à l'utilisation seule de l'encodeur).



Conclusion et perspectives :

Cette étude réalise un *benchmark* de compression de maillages hexaédriques hétérogènes issus des géosciences. La décomposition multi-échelles que propose HexaShrink devient un outil de compression de maillages efficace, associée à des encodeurs génériques sans perte récents, avec des taux de compression significatifs. Cependant, l'étude détaillée des performances sur chaque composante des maillages révèle une disparité dans la compressibilité. L'efficacité d'un encodeur générique peut en effet être impactée par la décomposition et la décorrélation des données. Cette observation oriente notre travail ultérieur vers l'usage d'encodeurs multi-échelles, plus à même d'exploiter les structures intrinsèques des données, et les corrélations entre différents coefficients [4]. Cette étude confirme donc l'intérêt multi-échelle de HS et ouvre de nouvelles pistes de recherche selon les propriétés.

Références (5 max.):

- [1] K.-A. Lie *et al.* Successful application of multiscale methods in a real reservoir simulator environment. Computat. Geosci., Dec. 2017.
- [2] J.-L. Peyrot *et al.* (H)exaShrink: Multiresolution compression of large structured hexahedral meshes with discontinuities in geosciences. IEEE ICIP, Sep. 2016.
- [3] J.-L. Peyrot *et al.* HexaShrink, an exact scalable framework for hexahedral meshes with attributes and discontinuities: multiresolution rendering and storage of geoscience models. Submitted, Sep. 2018.
- [4] P. N. Topiwala, editor. Wavelet image and video compression. Kluwer Academic, 1998.
- [5] A. Said and W. A. Pearlman. A new, fast, and efficient image codec based on set partitioning in hierarchical trees. IEEE Trans. CSVT, Jun. 1996.



Titre :

Tree-Structured Point-Lattice Vector Quantization for 3-D Point Cloud Geometry Compression

Auteurs :

Amira Filali, Vincent Ricordel, Nicolas Normand
LS2N Laboratoire, Polytech Nantes, rue Christian Pauc BP 50609, 44306 Nantes Cedex, France
{amira.filali, vincent.ricordel, nicolas.normand}@univ-nantes.fr

Résumé (100-200 mots) :

This paper deals with the current trends of technologies which aim at representing efficiently 3D point cloud, they have become a popular challenge for processing, storing and conveying the data independently of how it was captured. Compressing attributes of 3D point clouds such as geometry, colours or normal directions remain challenging problem, since these signals are unstructured. We propose an adaptive Tree-Structured Point-Lattice Vector Quantization (TSPLVQ), resulting in hierarchically structured 3D content, to improve compression performance for static point cloud. The novelty of the proposed approach lies in adaptively selecting the optimal quantization scheme for the point cloud geometric data, such that its intrinsic correlations can be better exploited. Two quantization modes are dedicated to project recursively the 3D point clouds into a series of embedded truncated cubic lattices. At each step of the process, the optimal quantization mode is chosen according to a rate-distortion criterion in order to achieve the best trade-off between coding rate and geometry distortion, such that the compression flexibility and performance can be greatly improved. To achieve rate-distortion optimization, the core of the proposed method relies on an accurate estimation of the distortion inside each cubic voronoï, between fitting plane and cloud points, taking into account the visual rendering of the 3D object. We experimentally evaluated the interest of the proposed method for geometry compression and visualization tasks.

Mots-clefs (3 à 5 max.) :

3D point cloud, Lattice Vector Quantization, Rate-Distortion Optimization, Compression

Contexte et état de l'art :

Advances in 360-degree video and virtual/augmented reality fields, provide many fast and reliable cameras, laser scanners, etc. which are rapidly spreading as devices for capturing the 3D data in different ways providing high definition 3D content. The highly increased quality and quantity of the captured data with additional reflectance and color produces very dense 3-D point cloud (PC) data to realistically represent the real world enabling immersive forms of interaction, navigation and communication. The common processing such as storage, transmission from the source data to the diverse platforms, or visualization, etc. become complex, or just impossible. In many industrial domains, it is desirable to transmit this data via restricted bandwidth networks. PCs provide a significant barrier for mass market applications. Thus, MPEG has issued a Call for Proposal (January 2017) for PC compression for its intended applications. Multiple methods of compression have been developed. In [1], Morell et al. have proposed an approach based on plane extraction which represents the points of each scene plane as a Delaunay triangulation and a set of points/area information. The concave hull of a plane is computed to find the borders and get a polygon representing it, an edge reduction is then performed to reduce the number of points needed. Many methods (available in the PCL [2]) are based on Octree decomposition to reduce the amount of data content. The principle is to build a large cuboid to contain all the PC data in a 3-D space. Then the algorithm splits the large cuboid into smaller 3D voxels of same size. For each voxel, the algorithm calculates the mean value of the 3-D coordinates of all the points. The mean value is used to represent all the points that lie in that voxel. In this way, the point cloud can be compressed while keeping the geometric details. In the context of classical images and video encoding applications, Ricordel et al. [4] have proposed a new vector quantizer (TSLVQ) based on truncated lattice embedding. They have investigated its complete design with: the lattice truncation, the multi-stage procedure of quantization, the unbalanced tree-structured codebook design according to a distortion v.s. rate trade-off. We propose to adapt the TSLVQ, the purpose of our method is to reduce the amount of data of a 3D points set but trying to preserve as much information as possible. To the best of our knowledge, we are the first, in the context of 3-D point clouds compression, to propose applying adaptive TSPLVQ with two quantization schemes (2x2x2 or 3x3x3 Voronoï splitting) depending on a rate-distortion criterion which takes into consideration the distortion between points and reconstructed planes associated to points in each cubic Voronoï taking into account the visual rendering of the 3D object and the tree coding cost.

Travail proposé (Décrire l'objectif du travail et indiquer clairement le problème/défi technique étudié) :

Tree-Structured LVQ (TSLVQ) is a gathering of many quantization approaches where the quantization is processed through a decision tree. TSLVQ [4] aims at using a hierarchical set of embedded lattices which is obtained such as it is possible to embed a lower scale truncated lattice into a cell of the next higher scale truncated lattice.

The use of the cubic lattice permits optimal embedding and fast quantization. The point cloud is highly complex and unstructured, such that an organization strategy of the point cloud can directly affect the coding performance. The proposed approach, based on the embedding of truncated lattices, permits hierarchical description of the 3-D point cloud through an adaptive tree-structure codebook simplifying its representation and compression. One single splitting mode, such as octree structure (where a cubic voronoï is splitted in 2x2x2), is insufficient to account for the characteristics of all kinds of point cloud geometries. Instead of a single splitting mode, in this work two splitting modes (2x2x2 versus 3x3x3) are employed to better map the point cloud to 3-D grids. We also have to scale and truncate the lattices suitably providing the best rate-distortion trade-off.

The basic quantization procedure used for our method TSPLVQ based on TSLVQ is illustrated in Fig. 2:

(0): Initialization: The points set is first normalized to fit inside the root voronoï cell by centering the points and then by scaling data with the factor: $F = \frac{1}{\sqrt{E_{max}}}$, where E_{max} is the maximal euclidean distance between a point and the center.

(1) and (2): An input vector x (its components are 3D point coordinates) inside a given cubic voronoï is scaled by the b factor and then shifted, the shifting U and factor b set up the splitting of the cube either in 2x2x2 ($b=2$, $U=(1/2,1/2,1/2)$) or in 3x3x3 ($b=3$, $U=(0,0,0)$).

(3): The fast quantization algorithm is then used to produce by rounding the corresponding reproduction vector Y .

(4): The output vector is centered to permit the next quantization level.

So to initialise the construction of the tree-structured codebook that will represent the 3D point cloud geometry, all the cloud points are projected into a first cubic Voronoï (namely the tree root). Next, the basic quantization structure is repeated iteratively from each point of the cloud. At each loop of the greedy process, the cube to split and the adapted splitting method, have to be chosen according to a rate-distortion criterion. Exactly a Lagrangian optimization is then employed at this level, the multiplier associated to each Voronoï (namely tree node) is computed, such as when a node is splitted, the increase in rate is calculated in terms of tree coding cost (here: the entropic cost of the nodes is considered), and the decrease in distortion in terms of geometric distortion. To assess objectively this distortion, we put in competition 2 metrics:

- Point-to-point metric, as the square distance between the lattice point Y_i and all assigned cloud points X inside the corresponding Voronoï cell C_i : $D_{P2Point}(X, Y_i) = \sum_{X \in C_i} d(X, Y_i)$

- Point-to-plane metric [3,5], as the square distance between the lattice point Y_i to normals $\vec{N}_{X_i}$ to the plane associated to the cloud points X inside the corresponding cell C_i : $D_{P2Plane}(X, Y_i) = \sum_{X \in C_i} ||\vec{V}_d(X, Y_i) \cdot \vec{N}_{X_i}||$

For a given loop of the tree growing process, the rate-distortion optimization is performed locally. It aims at choosing among all the possible splitted nodes, the one which offers the maximal decrease in distortion, and the minimal increase in rate. The codebook process is stopped for instance when the bit budget is reached, and the final tree-structure is then unbalanced. Each occupied leaf node corresponds to a point output. The location of the output representant is set to the average value of the PC points contained within the leaf node.

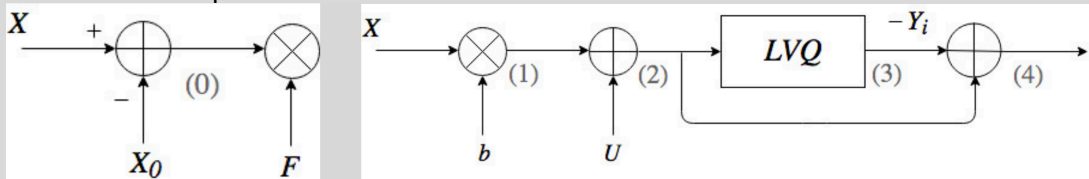


Fig.2. Initialization (left), basic procedure of TSPLVQ where X is the input vector, Y the reproduction vector, b the scaling factor (right).

Résultats : perspectives :

For the experimental results, we compared different voronoï splitting strategies: only 2x2x2 splitting mode (octree), only 3x3x3 mode, 2 splitting modes (2x2x2 vs. 3x3x3) have been put in competition. The performance comparison between the 3 different schemes are shown in Fig.3. Note that the distortion is assessed according to the 2 metrics that are considered: P2Point or P2Plane. For a fair comparison, we set up a parameter to our method, such as the rate in function of leaves number. It can be observed that the splitting strategies leads to different rate-distortion performan-

ce, for example, the 2x2x2 splitting scheme is more progressive and requiring more loops for the same number of obtained leaves compared to the 2 other schemes. The slopes of the curve is a key-point because when it becomes flat, it indicates that further splittings are not necessary. We can note that the slopes behaviors depend on the splitting method and of the used metric, and again the 2x2x2 method seems more interesting than the 3x3x3 method. When we put in competition the 2 splitting methods (called adaptive method), the metric choice is important: with P2Point the 2 splitting methods are combined and the slopes are not flat, with P2Plane no combination are made. At this level of the work, it also seems that the P2Plane metric does not show advantage in term of objective quality evaluation. But it is also important to consider the rendering of the encoded point cloud, Fig.4 presents the rendering coded 3D point cloud in order to analyze and compare their visual quality. Our adaptive method preserves more visual details in the complex areas (for instance the head of the dragon) compared to the two other splitting schemes, whereas it produces less points to represent smooth surfaces (for instance the dragon body surface). Hence our proposed adaptive method using P2Plane distortion seems more linked with the subjective quality of the rendered decoded object.

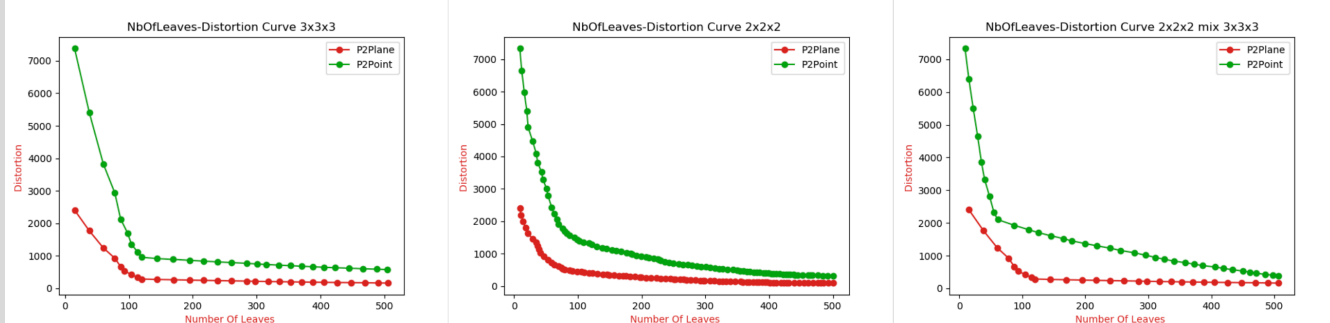


Fig.3. TSPLVQ method using P2Point vs. P2Plane distortion metric Of « Dragon.ply ».

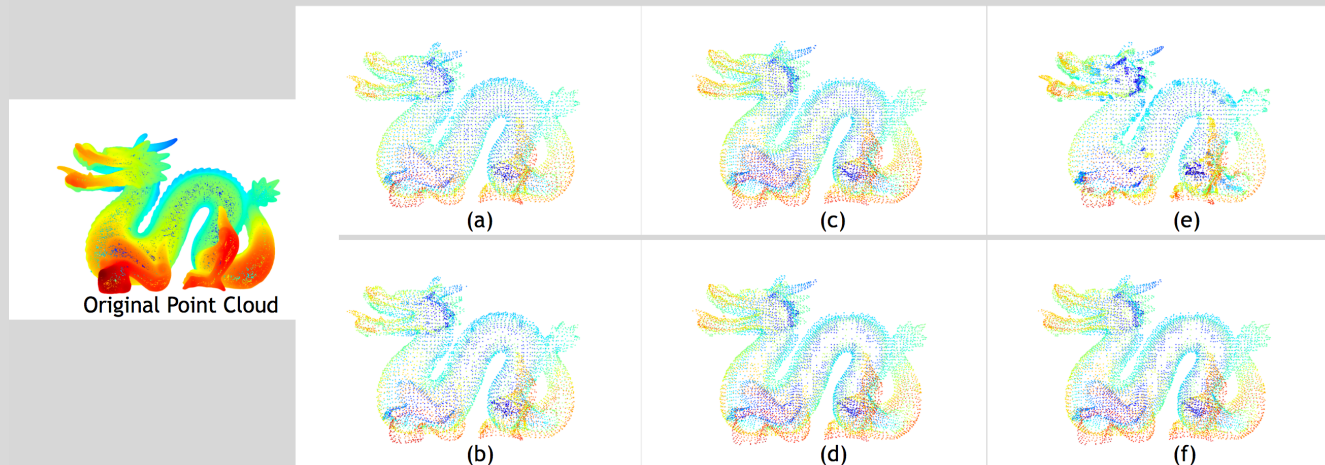


Fig.4. Visual rendering result of Dragon point cloud: (a), (b) only 2x2x2 splitting using P2Point and P2Plane distortion respectively; (c), (d) only 3x3x3 splitting using P2Point and P2Plane distortion respectively; (e), (f) 2x2x2 vs. 3x3x3 splitting using P2Point and P2Plane distortion respectively.

Conclusion et perspectives :

Our main research investigated into a compression method of the geometric data of PCs. We have proposed a lossy compression method based on TSPLVQ. Our method innovates because it introduces three splitting strategies (2x2x2 versus 3x3x3) for the cubic voronoï, and by the use of two metrics (P2Point vs. P2Plane). Experimental results show the different behaviours of the rate-distortion curves and different qualities of the rendered contents, their analysis explains the TPSLVQ coding improvement in this context. As future work, we would like to include more accurate geometric primitives and other attributes (as colour) to better characterise the content of a voronoï cell, in order to better render the decoded 3D objects. We also plan to improve the rate computation for the coding of the tree-structure.

Références (5 max.) :

- [1] Morell, V., Orts, et al., "Geometric 3D point cloud compression", Pattern Recognition Letters (2014).
- [2] Rusu, R. B., Cousins, S., "3D is here: Point Cloud Library (PCL)", International Conference on Robotics and Automation (2011).
- [3] Podulka, P.: "selection of reference plane by the least squares fitting methods", Advances in Science and Technology Research Journal (2016).
- [4] Ricordel, V. and Labit, C., "Vector Quantization by packing of embedded truncated lattices", ICIP (1995), Washington.
- [5] Alireza, J., et al., "Subjective and Objective Quality Evaluation of Compressed Point Clouds", 19th International Workshop on MMSP (2017).



Titre :

DNA CODING FOR IMAGE STORAGE USING IMAGE COMPRESSION TECHNIQUES

Auteurs :

Melpomeni Dimopoulou ¹ dimopoulou@i3s.unice.fr	Marc Antonini ¹ am@i3s.unice.fr	Pascal Barbry ² barbry@ipmc.cnrs.fr	Raja Appuswamy ³ Raja.Appuswamy@eurecom.fr
--	---	---	--

¹ I3S, 2000, Route des Lucioles, 06900, Sophia Antipolis, France

² IPMC, 660 Route des Lucioles, 06560, Sophia Antipolis, France

³ EURECOM, 450 Route des Chappes, CS 50193 - 06904 Sophia Antipolis, France

Résumé (100-200 mots) :

Living in the age of the digital media explosion the urge for finding new efficient methods of data storage increases significantly. Existing storage devices such as hard disks, flash, tape or even optical storage have limited durability in the range of 5 to 20 years. Recent studies have proven that the method of DNA data storage introduces a strong candidate to achieve data longevity. The DNA's biological properties permit the compression of a great amount of information into an extraordinary small volume while also promising efficient data storage for millions of years with no loss of information. This work proposes a new encoding scheme especially designed for the encoding of still images, extending the existing algorithms of DNA data storage by introducing image compression techniques.

Mots-clefs (3 à 5 max.) :

DNA data storage, wavelet decomposition, data compression, error correction

Contexte et état de l'art :

DNA data storage is the procedure of encoding any binary information into a quaternary code of A, T, C, G, the symbols that correspond to the 4 different types of nucleotides (nts), the DNA's main building blocks. The selection of the encoding algorithm is strongly restricted by the biological procedures involved in the encoding scheme in figure [1]. DNA synthesis is the procedure of chemically producing chunks of nucleotides (oligos) corresponding to the desired data. This process is almost error-free with an error probability lower than 0.1% for DNA sequences no longer than 100 nts. The DNA sequencing is the procedure of reading the stored oligos and recover the initial quaternary encoded sequence that has been formed into DNA. However, the biological procedure of sequencing is a significantly error-prone procedure. G. Church et al. in [1] have studied the main causes of the biological error while in [2] Goldman et. al proposed an efficient encoding procedure to prevent this kind of errors as much as possible and introduced error correction. In addition to error-correction algorithms, sequencing machines use the method of PCR amplification creating many copies of each oligo introducing redundancy to reduce the biological error as much as possible. One can imagine the procedure of PCR amplification like the classical repetition coding used for transmission over a noisy channel that may corrupt the transmission in various positions. As in repetition coding, the main idea of PCR amplification is to repeat the oligos several times hoping that the sequencing corrupts only a minority of these repetitions. Under this simple hypothesis, we assume that after all the copies of oligos are sequenced, one can apply the method of majority vote to distinguish the most representative oligos. In other words, under the main assumption that the oligos that appear more times than the others will most probably be error-free, during the reconstruction we select the oligos with the highest frequency of appearance.

Travail proposé (Décrire l'objectif du travail et indiquer clairement le problème/défi technique étudié) :

In this work we have built an encoder specifically designed according to the needs of the biological procedures of synthesis and sequencing. First, we compress the input image using image compression techniques. More precisely, we use the Discrete Wavelet Transform (DWT) to decompose the image into 3 levels producing 10 wavelet subbands. Then we encode each subband separately using uniform quantization for compression. As a next step we need to encode the quantized coefficients into a quaternary code of A, T, C, G respecting four

basic restrictions imposed by the biological procedures. To begin with, the sequence should not contain repetitions of the same nucleotide more than 3 times (homopolymers). Additionally, the encoded sequence should not contain repetition of the same codewords and the percentage of G, and C should be lower or equal to the percentage of A and T. This means that the succession of symbols should be as random as possible so to prevent errors. Finally, as the synthesis error increases exponentially with the increase in the sequence length, the encoded sequence to be sent for synthesis should not exceed the 100nts. This last restriction imposes the need for cutting the encoded subband sequence into smaller chunks of 91 nts and adding a header section in each one of them that will contain information about their order in the initial sequence. After the encoding procedure the encoded oligos are biologically synthesized into synthetic DNA which is stored in a small tube depicted in image [1]. In the decoding part the stored oligos are being replicated in many copies before they are read with the sequencing procedure which further adds sequencing noise. As a last step we choose the most representative oligos and we reconstruct the initial image following the inverse procedure. In this first experiment we tested two different cases of reconstruction. One by using the most frequent oligos and one choosing the oligos randomly. This experiment produced 662 oligos compressing the original image at a bit rate of 1.96 nts/pixel (equivalently 4 bits/nt) and a PSNR of 32.51 dB. The results are presented in figures [2] and [3] respectively.

Résultats :

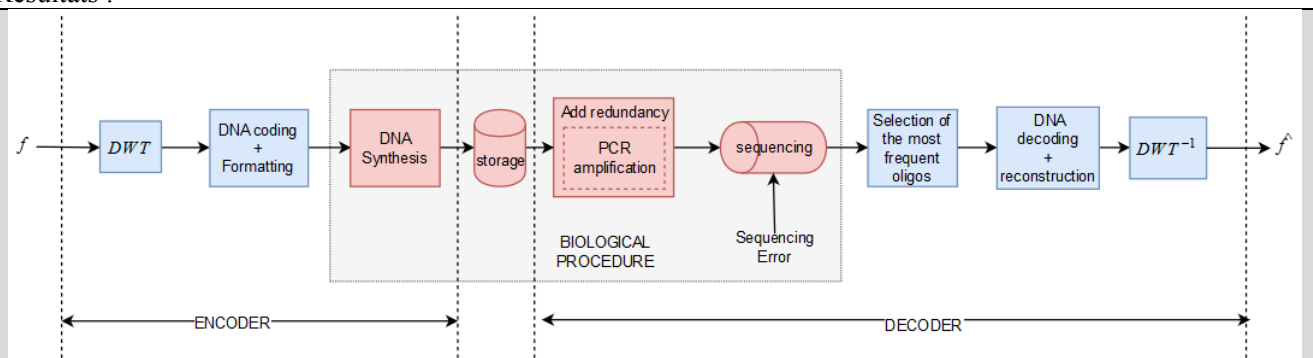


Figure 1: The encoding scheme. The initial image is being decomposed into wavelet subbands and each one of them is encoded into a quaternary code and cut into smaller chunks. The encoded sequence is synthesized into DNA and stored in a tube. PCR amplification produces many copies of each oligo creating redundancy. The copies are being read through the sequencer which introduces sequencing error. The most representative oligos are then selected to decode and reconstruct the encoded sequence and using inverse DWT an estimation of the initial image is produced.



Figure 2 (left): The reconstructed image using the most frequent oligos. (No sequencing noise)



Figure 3 (left): The reconstructed image using random selection of oligos.



Image 1 (left): The storage tube containing 103 ng of synthetic DNA

Conclusion et perspectives :

Given the results of figures [2] and [3] one can clearly conclude that choosing the most frequent oligos from the different copies provided by the sequencing, it is more probable to achieve the best possible reconstruction of the image. In this experiment we have managed to get a reconstruction without any sequencing noise. This very first and simple attempt to introduce image compression techniques into DNA data storage has provided very promising results. Future experiments using error correction techniques and introducing more robustness to sequence error, will further improve the compression ratio and the quality of the current results.

Références (5 max.) :

- [1] Church, G.M., Gao, Y. and Kosuri, S., 2012. Next-generation digital information storage in DNA. *Science*, p.1226355.
- [2] Goldman, N., Bertone, P., Chen, S., Dessimoz, C., LeProust, E.M., Sipos, B. and Birney, E., 2013. Towards practical, high-capacity, low-maintenance information storage in synthesized DNA. *Nature*, 494(7435), p.77.



Titre :

A variant of Local Binary Pattern applicated on ear recognition in unconstrained condition

Auteurs :

Youbi Zineb⁽¹⁾, Larbi Boubchir⁽²⁾, Abdelhani Boukrouche⁽¹⁾

⁽¹⁾ PI:MIS Laboratory, University of 8 Mai 1945 Guelma, Algeria

⁽²⁾ LIASD research Lab., University of Paris 8, France

Résumé (100-200 mots) :

The recognition of people from their ear images is a research focus within the biometric community. The ear is stable over time and it is rich in features such as texture, color and size.

In this article, we propose an ear recognition system, based on a variant of local binary pattern named multi scale local binary pattern MLBP. The main idea is summed up by dividing the image over several overlapping regions and applying the MLBP on each zone. the concatenation of the histograms resulting from the zones forms the features vector.

We have evaluated the proposed method on three reference ear databases: IIT Delhi I, IIT Delhi II and USTB-1. The results obtained have clearly shown that the proposed method is robust, stable feel and very competitive, it achieved an attractive recognition performance in terms of identification rate at rank-1 up to: 97%.

Mots-clefs (3 à 5 max.) :

Ear biometrics, identification, feature extraction.

Contexte et état de l'art :

Recently, there is an emergence of utilization of the human ear as a biometric trait. The richness and stability of the structure are the characterization of the human ear pattern which provides a large amount of information such as size, color and texture; allowing to differentiate between people. The ears appearance do not influence with the face expressions and age, it is a non-intrusive modality it do not need a big cooperation with the captor. Although the effect of gravity, ear accessories, hair and earphone can perturb the length of the ear lobe. Table 1 shows a comparison between the best identification performances of the state of art methods.

Authors	Method	Results
Basit and Shoaib (2014)	Non linear curvelet features	97.77 %
Jacob and Raju (2014)	Gray level co-occurrence matrix	94.12 %
Benzaoui et al. (2014)	Besif descriptor with KNN descriptor	97.26 %
Yobui et al. (2016)	Local texture descriptors	95.02%
Benzaoui et al. (2017)	Local texture descriptors	97.39 %
Mamta and Madasu (2013)	Traditional Principal Component Analysis	79 %

Table 1. Comparing the best identification performance of the state-of-the-art methods [1]

Travail proposé (Décrire l'objectif du travail et indiquer clairement le problème/défi technique étudié) :

Recently, the use of ear as a biometric trait is emerged. Its structure is reach and stable, and it provides a characteristic (such as size, colour and texture) which allowing to differentiate between people. It do not influence with the face expressions and age, it is a non-intrusive modality. Although, a perturbation is possible

with the effect of gravity, ear accessories, hair and earphone [1].

The goal of our work is to provide a reliable ear recognition system, under uncontrolled conditions. The proposed method is constituting of two major steps, feature extraction and matching process. Firstly, image is pre-processed and the features are extracted with Multi scale local Binary Pattern [2], by labeling the pixels of the image by thresholding a block of the neighborhood of each pixel (i, j) with the center value and summing the thresholded values weighted by powers of two. It is computed in a local circular region (P, R) where R is the radius of the circular neighbourhood and P is the number of the neighbours.

$$LBP_{P,R}(i, j) = \sum_{p=0}^{P-1} 2^p S(X(i, j) - X(i_p, j_p))$$

The occurrences of LBP codes of an image are collected into a histogram H , is used as texture feature vector

$$H(n) = \sum_{i,j} B(LBP_{P,R}(i, j) = n) \text{ where } n \in [0, 2^P - 1]$$

$$B(v) = \begin{cases} 1, & \text{when } v \text{ is true} \\ 0, & \text{otherwise} \end{cases}$$

MLBP features are extracted from the ear image and then combined by concatenating the set of histograms H computed at different scales. In our proposed method, we propose to divide the image into equal blocs and then we compute MLBP on each bloc. these features are stored in the database and called template. Secondly, the ear image was also pre-processed and the features are extracted. These features are then compared using a matching process against the template stored previously in the template to determine or to verify the identity of the person.

The following algorithm summarizes our proposed human ear recognition method.

Algorithm 1 Human ear recognition based on MLBP and CTB distance

Input : Ear image X

- 1- convert X to 256 levels of gray
- 2- filter X with median filter
- 3- Divide X into K equal blocks: $B_{(k)}$; $k=1, \dots, K$
- 4- **for** $k = 1$ to K **do**
- 5- Compute MLBP on the bloc $B_{(k)}$
- 6- **endfor**
- 7- Concatenate the computed MLBP features: $\mathbf{F} = \{H_{(1)}, \dots, H_{(K)}\}$
- 8- Apply CTB distance

Output: Identification decision

Résultats :

We have evaluated the proposed method on three reference ear databases: IIT Delhi I [3], IIT Delhi II [3], USTB-1 [4]. Each subject of these databases has at least three ear images. We have taken for each subject, two images in train and the rest for test. We have switched the position of the training images between the first three images and then the average rate of the three cases is taken as an identification rate. To evaluate the performance of the proposed approach under pose variations, we tested its robustness against several changes in rotation. We randomly rotated each test image by an angle θ in the interval $[-10 \text{ deg}; +10 \text{ deg}]$. As the size of the entered image is changed by the effect of rotation, the size of the feature vector will also be automatically changed. We, therefore, selected the centered region ROI from the rotated image. This ROI must have the same size as the entered image to generate a feature vector that has the same size as the vectors stored in the database. The rank-1 recognition rates of the IIT Delhi-1, IIT Delhi-2, and USTB-1 databases using two referential ear images per person under different angles of θ up to: 97%.

Conclusion et perspectives :

In this work, we have proposed a novel approach for automatic human ear biometric recognition using local texture features based on MLBP descriptor in uncontrolled conditions. The results obtained using IIT Delhi1, IIT Delhi 2, and USTB-1 databases have shown that the proposed method achieves an improved performance

for identification mode up to 97%; and outperforms the state of-the-art. The investigation of other multi-scale representations; for feature extraction improvement is expected to be our forthcoming work.

Références (5 max.):

- [1] Z. Youbi, L. Boubchir, M. Bounneche, A. Ali-Cherif, A. Boukrouche, "Human ear recognition based on multi-scale local binary pattern descriptor and KL divergence", in the Proc. of the IEEE 39th Int. Conf. on Telecommunications and Signal Processing (TSP), pp.685–688, 2016.
- [2] M. Pietikäinen, A. Hadid, G. Zhao, T. Ahonen, "Computer Vision Using Local Binary Patterns", Springer-Verlag London Limited, 2011.
- [3] A. Kumar and C. Wu, "Automated human identification using ear imaging", Pattern Recognition, vol. 45, no. 3, pp. 956-968, 2012.
- [4] Z. Mu, "USTB Ear Image Database Beijing", <http://www1.ustb.edu.cn/resb/en/index.htm>, 2009.



Titre :

Algorithme d'appariement de blocs basé sur l'optimisation de la compensation de disparité pour le codage d'image stéréoscopique

Auteurs :

Imen KADRI⁺⁺, Gabriel DAUPHIN⁺, Anissa MOKRAOUI⁺, Zied LACHIRI⁻

⁺L2TI, Institut Galilée, Université Paris 13 Sorbonne Paris Cité

99, avenue Jean-Baptiste Clément 93430 Villetaneuse, France

⁻LSITI, Ecole Nationale d'Ingénieurs de Tunis, Université de Tunis El Manar

BP 37, Campus Universitaire, Le Belvédère, 1002, Tunis, Tunisie

Résumé :

Cet article présente un algorithme d'appariement de blocs basé sur la compensation optimisée de disparité pour le codage d'image stéréoscopique. Notre algorithme propose d'estimer une première carte de disparité basé sur l'algorithme classique de mise en correspondance des blocs similaires dans les deux vues en minimisant l'erreur quadratique moyenne entre la vue originale et sa version reconstruite sans compensation de disparité. Cette carte est ensuite modifiée de façon à ce que la vue prédite puis compensée donne de meilleures performances en termes de débit-distorsion comparées au schéma classique de codage. Pour un bloc donné, l'algorithme propose de tester l'ensemble des disparités de la fenêtre d'appariement choisie en tenant compte de l'impact de l'erreur résiduelle décodée sur la vue reconstruite. Parmi les disparités testées, l'algorithme sélectionne alors celle qui améliore la vue reconstruite après compensation pour un débit cible. Basé sur ce principe, les disparités de la carte initiale sont toutes mises à jour. Les simulations montrent que notre algorithme non seulement réduit la redondance inter-vue mais également améliore la qualité de la vue reconstruite et compensée par rapport à la méthode classique de codage avec compensation de disparité.

Mots-clefs :

Image stéréoscopique, mise en correspondance de blocs, disparité, compensation, codage

1. Contexte et état de l'art :

Une image stéréoscopique est composée de deux vues acquises selon deux points de vue légèrement différents de la même scène. De ce fait, les images stéréoscopiques nécessitent deux fois plus d'information comparée à une image 2D. Il s'avère alors nécessaire de se préoccuper de la question du codage de ces images stéréoscopiques pour les stocker ou les transmettre. Dans cet objectif, de nombreux travaux se sont intéressés à la réduction de la redondance inter-vue [1]. L'idée sous-jacente étant de reconstruire l'une des deux vues en utilisant la deuxième vue (vue de référence) et la carte de disparité associée à la vue à reconstruire. En effet la carte de disparité est moins coûteuse en termes binaire que le codage d'une des deux vues. Cette dernière, associée à l'une des deux vues, est généralement estimée par l'algorithme classique d'appariement de blocs similaires dans les deux vues. Notons que la qualité de la vue reconstruite dépend fortement de la carte de disparité estimée. De ce fait, pour tenir compte des erreurs de reconstructions trop importantes dues par exemple aux changements de luminosité, aux zones texturées ou aux occultations, la majorité des travaux proposent des solutions de codage basées sur le principe de la compensation de disparité. L'objectif de cet article est d'améliorer les performances en termes de débit-distorsion de ce schéma classique de codage.

2. Algorithme d'appariement de blocs basé sur l'optimisation de la compensation de disparité :

Dans cet objectif, nous proposons d'estimer la carte de disparité par blocs (notée $\{d(k, l, Q)\}$) qui minimise l'erreur quadratique moyenne entre l'image originale de la vue droite $I_d(i, j)$ et sa version prédite compensée par l'image résiduelle décodée notée $\hat{I}_d(i, j) = I_c(i, j, s, Q)$. L'image résiduelle, donnée par $(I_d(i, j) - \hat{I}_g(i, j + d(k, l, Q)))$ avec $\hat{I}_g$ l'image gauche codée, est codée par JPEG selon le facteur de qualité Q choisi. L'estimation de la carte de disparité est alors formalisée par l'équation suivante :

$$d(k, l, Q) = \underset{s \in S}{\operatorname{argmin}} \left(\sum_{(i,j) \in B_{k,l}} (I_c(i, j, s, Q) - I_d(i, j))^2 \right),$$

où S représente la fenêtre d'appariement choisie, $B_{k,l}$ est le bloc à apparier de coordonnées (k, l) .

Pour résoudre cette équation, nous proposons la démarche brièvement résumée ci-dessous qui consiste à estimer une première carte de disparité. Cette dernière cherche à minimiser l'erreur quadratique moyenne entre l'image originale associée à la vue droite et sa version uniquement prédite $\hat{I}_d(i, j) = I_g(i, j + s)$:

$$d(k, l) = \underset{s \in S}{\operatorname{argmin}} \left(\sum_{(i,j) \in B_{k,l}} (I_g(i, j + s) - I_d(i, j))^2 \right).$$

L'ensemble des disparités de cette carte initiale sont mises à jour, une à une, en testant toutes les disparités possibles de la fenêtre d'appariement et plusieurs facteurs de qualité Q pour comprimer l'image résiduelle. Pour une disparité choisie dans S , l'algorithme opère comme suit : (i) l'image de la vue droite est reconstruite ; (ii) l'image résiduelle est à son tour codée selon JPEG en utilisant plusieurs facteurs de qualité Q ; (iii) pour chaque facteur de qualité, l'image résiduelle est décodée pour compenser l'image prédite. Les performances sont mesurées en termes de débit-distorsion. Le débit binaire tient compte du coût de codage entropique de la carte de disparité ainsi que le coût de codage du résiduel associé à l'image droite I_d . La distorsion s'appuie sur le PSNR où l'erreur quadratique moyenne est calculée entre l'image de la vue droite originale I_d et sa version prédite compensée par le résiduel décodée. Cette procédure est répétée pour toutes les valeurs possibles de la fenêtre d'appariement. On retiendra la disparité qui donnera le meilleur PSNR pour le débit cible. Ce processus est itéré tant que les disparités de la carte initiale n'ont pas toutes été mises à jour.

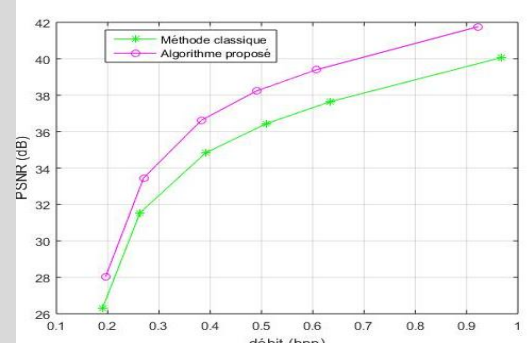
3. Résultats de simulation :

Dans les simulations réalisées, le décodeur dispose de l'image gauche originale. Les tests sont réalisés sur les images extraites de la base Middlebury [2] et Demos. Le débit binaire tient compte du coût de codage de la carte de disparité (codage arithmétique) et de l'image résiduelle par JPEG. La taille des blocs est fixée à 8×8 . Les performances de notre algorithme sont comparées à celles d'un schéma de codage classique. Le tableau ci-dessous présente les performances obtenues sur l'image droite. La métrique Bjontegaard montre que la valeur moyenne du gain en PSNR à débit donné de l'algorithme proposé par rapport au schéma classique est en faveur de notre algorithme. Nous observons que notre algorithme permet une très nette amélioration pour certaines images. Les courbes débit-distorsion de la Figure 1 montrent un gain en termes de PSNR relativement constant sur un large intervalle de valeurs de débit (entre 0.4 bpp et 1 bpp).

Tableau 1 :
Métrique
Bjontegaard

Image	Δ PSNR (dB)	bpp (%)
Bowling	+1,85	-20,8 %
Books	+1,65	-27,5 %
Teddy	+1.1	-21 %
Barn	+0.44	-13 %
Voiture	+0,1	-4,4 %

Figure 2 :
Comparaison des
performances
en termes de
débit-
distorsion.



4. Conclusion et perspectives :

Les bonnes performances de l'algorithme proposé dans cet article montrent bien l'intérêt d'estimer les disparités en tenant compte de la prédiction après compensation effective. La grande complexité numérique de cet algorithme nous amène à réfléchir sur l'utilisation de techniques d'apprentissage pour estimer plus simplement les disparités tout en maintenant les bonnes performances avec un temps de calcul relativement raisonnable.

Référence :

- [1] F. Dufaux, B. Pesquet-Popescu, M. Cagnazzo: Emerging Technologies for 3D Video: creation, coding, transmission and rendering, 2013, John Wiley & Sons.
- [2] <http://vision.middlebury.edu/stereo/data/>



Titre :

Clustering Algorithm Based on Shuffled Frog Leaping Algorithm

Auteurs :

Sami Jaballah, Mohamed-Chaker Larabi
XLIM UMR CNRS 7252, University of Poitiers, France

Résumé (100-200 mots) :

Cluster analysis aims to determine a set of patterns into a partition or so-called clusters based on similarity their statistical properties. The main goal is to maximize the homogeneity between data points within a cluster and the heterogeneity between different clusters. The proposed clustering method has been evaluated on 2 artificial datasets and five real-life datasets and has been compared with four state-of-the-art clustering methods. Experimental results show that the proposed algorithm achieve better clustering performance in term of sum of intra-clusters distances.

Mots-clefs (3 à 5 max.) :

Clustering, shuffled frog leaping algorithm, machine learning

Contexte et état de l'art :

Clustering is the task of partitioning data points into k groups or clusters, while ensuring similarity within the cluster and the dissimilarity with the other clusters. K-means is one of the most widely used center-based algorithm thanks to its effectiveness for many data clustering problems. However, K-means suffers from some flaws. The final clusters of the K-means are sensitive to the initial seed and its objective function is multi-modal shaped that can be easily trapped into a local optimum which will result into an empty cluster. To overcome these drawbacks many authors have proposed the use of evolutionary algorithms in the clustering analysis domain [1]. These algorithms have the ability to provide a near-optimal solution to the NP-hard problems with a large number of variables and non-linear objective functions [2]. In order to enhance the efficiency of the clustering, we opted for the use of the improved shuffled frog leaping algorithm (ISFLA) [3] to reduce the sum of intra-cluster distance. The proposed method is compared to K-means, particle swarm optimization (PSO) [4], and the original SFLA.

Travail proposé (Décrire l'objectif du travail et indiquer clairement le problème/défi technique étudié) :

The steps of the ISFLA clustering algorithm are described as follows:

1. Frog's representation: In the context of clustering, each Frog (individual) contains K center vectors.
2. Random Frog: The position of the frog F_i is calculated as follows:
 - Assign K different random feature vectors from the entire data set to the frog position $x F_{id}^R$
 - Randomize the resulted position as $x F_{id} = x F_{id}^R \cdot rand()$.
3. Population initialization: generate N random frogs as described in step 2.
4. Fitness function: the fitness value of each Frog is the sum of intra-cluste distances, i.e. the distances between each cluster center and assigned data vectors is defined as:

$$F = \sum_{i=1}^N \text{Min} \{d_{ij} (X_i, V_j)\}$$

where d is the Euclidian distance and N is the total number of data vectors.

Résultats :

Table 1: Result of the clustering performance (F) obtained by the four algorithms for 50 runs over 7 datasets.

Data set	Criteria	K-means	PSO	SFLA	ISFLA (proposed)
Art1	Best	518.77	519,94	523,09	518,56
	Average	529.93	523,83	526,84	518,58
	Std	49.67	1,96	2,83	0,01
Art2	Best	1676.96	1675,34	2765,13	1675,32
	Average	2126.385	1926,65	3442,07	2293,29
	Std	378.6	431,05	272,36	445,2453
Vowel	Best	149380,28	149062,07	233103,44	148967,24
	Average	152780,80	158894,59	256067,98	168150,11
	Std	4305,38	6895,11	12823,70	18467,26
Iris	Best	97,32	96.66	130,95	96,65
	Average	102,48	105,95	145,48	105,61
	Std	10,56	14,58	9,24	14,11
CMC	Best	5542,18	5532,18	6079,47	5532,18
	Average	5543,36	5532,18	6451,57	5532,18
	Std	1,45	5,74604E-9	155,81	5,79E-12
Crude	Best	279,271	277,51	292,3853	277,21
	Average	279,68	278,18	308,0043	277,22
	Std	0.13	0,43	9,35421	0,03
Wine	Best	16555,67	16292,18	16607,10	16292,18
	Average	16743,80	16292,67	16891,45	16292,28
	Std	579,04	0,68	169,95	0,19

To reduce the effect of randomness, 50 runs are performed for each dataset and three parameters were extracted: the optimal solution, the average and the standard deviation. As shown in Table 1, the proposed method has the lower standard deviation values for most of the tested datasets and its average values are mostly close to the best solution values. We can conclude that the ISFLA clustering algorithm is robust to the random initial population and it converges consistently to the best solution.

Conclusion et perspectives:

In this paper, a novel clustering technique based on the improved shuffled frog leaping algorithm is proposed. Experiments were carried out on 2 artificial datasets and five real-life datasets in comparison with other clustering techniques: K-means, PSO, SFLA. The results illustrate that the ISFLA outperforms the other algorithms and converges to a near-optimal solution for clustering problem.

Références (5 max.):

- [1] Cole, Rowena Marie. *Clustering with genetic algorithms*. University of Western Australia, 1998.
- [2] Das, Swagatam, Ajith Abraham, and Amit Konar. "Automatic clustering using an improved differential evolution algorithm." *IEEE Transactions on systems, man, and cybernetics-Part A: Systems and Humans* 38.1 (2008): 218-237.
- [3] Jaballah, Sami, et al. "An improved shuffled frog leaping algorithm with a fast search strategy for optimization problems." *Intelligent computer communication and processing (ICCP), 2014 IEEE international conference on*. IEEE, 2014.
- [4] Eberhart, R. C., and J. Kennedy. "Particle swarm optimization, proceeding of IEEE International Conference on Neural Network." *Perth, Australia* (1995): 1942-1948.



Titre :

Robust Image Segmentation for Early Plant Diseases Detection on Leaf

Auteurs :

Arnaud S. M. R. Ahouandjinou and Cina Motamed

Résumé (100-200 mots) :

Image segmentation of plant leaf is an important tool in the early detection of plant diseases from symptoms at the parasite control stage during flowering and maturity. Indeed, in the phase of pest control in order to resist plant diseases in agriculture, the current solution requires the massive use of phytosanitary products that are both dangerous for crops, the agricultural environment, and the health of both farmers and consumers. The main goal of this work is to respond to a new agronomic challenge in the context of agro-ecology through the development of a new method of pest control by using image processing tools. For this, we propose a novel region-based segmentation approach in a credibility context. This approach has the advantage to take into account the segmentation model of all the sources of noise during the acquisition step. The experimental results and comparison results with similar approaches demonstrated that the proposed method is efficient and very robust to the noises of the image acquired by its genericity.

Mots-clefs (3 à 5 max.) :

Plant disease detection on leaf, image processing, region based-segmentation, Transferable Belief Model (TBM)

Contexte et état de l'art :

Image segmentation is a fundamental yet still challenging problem in computer vision and image processing. In particular, it is an essential process for many applications such as object recognition, object detection, target tracking, content-based image retrieval and medical image processing, etc. But also increasingly exploited in recent years in the agricultural field to the advent of smart system for crop monitoring. In agriculture research of automatic leaf disease detection is essential research topic as it may prove benefits in monitoring large fields of crops, and thus automatically detect symptoms of disease as soon as they appear on plant leaves. Bare eye based plant disease detection is not efficient so, it is useful to make recourse for image processing techniques are used. In addition, it becomes even more difficult with a biased and therefore less accurate estimate of a large agricultural area. Furthermore, Phytosanitary products used in agriculture to fight against plant diseases are molecules foreign to biological processes and leave residues in the environment. They have real consequences for the health of users and consumers because of their persistence. Therefore, it is useful to use image processing techniques such as image segmentation, which is already widely used in literature reviews for this type of application. Nevertheless, the main limitation of this technique is its poor noise resistance and its accuracy which can be improved for critical applications. There is little disease detection related work nevertheless from plant leaves of a state analysis of the no less important art works and techniques used is made in this section. Indeed, disease management is a challenging task in agriculture. Mostly diseases are seen on the leaves or stems of the plant. Precise quantification of these visually observed diseases, pests, traits has not studied yet because of the complexity of visual patterns [2]. Hence there has been increasing demand for more specific and sophisticated image pattern understanding [3]. We have various Types of leaf spot diseases such as bacterial, fungal and viral. K. Thangadurai and K. Padmavathi [1] recommended computer vision image enhancement for leaf disease identification [x. Mrunalini R. B. et al proposed on two techniques for feature extraction [2]. In [3], the authors proposed an image segmentation approach based on edge detection is performed, and then image analysis and disease classification are performed using homogeneous cotton disease detection algorithm.

Travail proposé (Décrire l'objectif du travail et indiquer clairement le problème/défi technique étudié) :

This work focused the design of robust technique for image processing that have been used for detection of plant diseases. The proposed approach is robust to variation in brightness, and more generally all other forms

of noise at the image analysis. At the first step, Images are captured by digital camera mobile networks and the second step, enables to process image using the region growing based approach segmentation in the evidential framework. This paper provides an efficient method used to make of leaf disease detection using image processing. The proposed method is to increase throughput and reduction subjectiveness arising from human experts in detecting the leaf disease [1]. The main contribution of this work is to promote healthy farming and precise agriculture through the reduction use of pesticides and better planning exploitation of these to minimize the risk of environmental pollution while improving the quality of agricultural products. To improve agricultural products, automatic detection of symptoms are beneficial in order to contribute to the reduction of hunger in the world and even more in the developing countries. Plant leaf analysis based on image processing using for instance segmentation technique can be applied for the following purposes: 1) To detect disease affected leaf, stem, fruit; 2) To quantify affected area by disease; 3) To find the boundaries of the affected area ; 4) To determine size & shape of leaf; 5)To identify the object correctly. Figure 1, shows the global proposed segmentation approach which is divided in three main parts such as: noise estimation, Extraction of image characteristics which are more relevant because of insensitive to acquisition noise, and the disease detection over image segmentation step.

Based on state-of-the-art techniques used for image segmentation of plant leaves, we can also notice a wide use of segmentation techniques based on thresholds and region. Comparative study of these methods are done using some standard parameters such as: spatial information, region continuity, speed, computation complexity, automaticity, noise resistance, multiple object detection and accuracy. Given the performance evaluation criteria of a segmentation algorithm, we find better performance for thresholding and region based segmentation even if these techniques are very sensitive to noise and the accuracy of the results are still to be improved. For this reason, in our approach, we have opted for these categories of methods but improving both sensitivity to noise as well as accuracy and performance. The results of the work in [7] inspired a lot the method we proposed. Our approach is based on the exploitation of three main parameters (grayscale, color and texture) as in [7] sufficient to describe the image fully. To facilitate the global management of its parameters, we have combined them in a single variable with prudent information fusion operators in TBM framework reasoning [8]. Moreover, in the model a sources of imperfections (variation of the brightness: morning, noon and evening) which is an acquisition noise is also modeled and integrated in the model to improve the sensibility and accuracy of the quality of the detection. In our queuing approach, we have three characteristics that are the gray level, the color and finally the texture. A phase of extraction essentially makes it possible to measure the level of gray, to characterize the image in the color space and finally to extract the texture [6]. For each pixel of the images, each of its parameter is estimated and transformed into a probabilistic variable. Then, its parameter values are combined into a mass in a Transferable Belief Model [7]. TBM is a general framework for uncertainty representation and combination of various pieces of information without additional priors. This mass is then transformed in communality in order to facilitate the decision of part of a pixel to a region. Figure 1 shows the global proposed segmentation approach which is divided in three main parts such as: noise estimation, Extraction of image characteristics which are more relevant because of insensitive to acquisition noise, and the disease detection over image segmentation step.

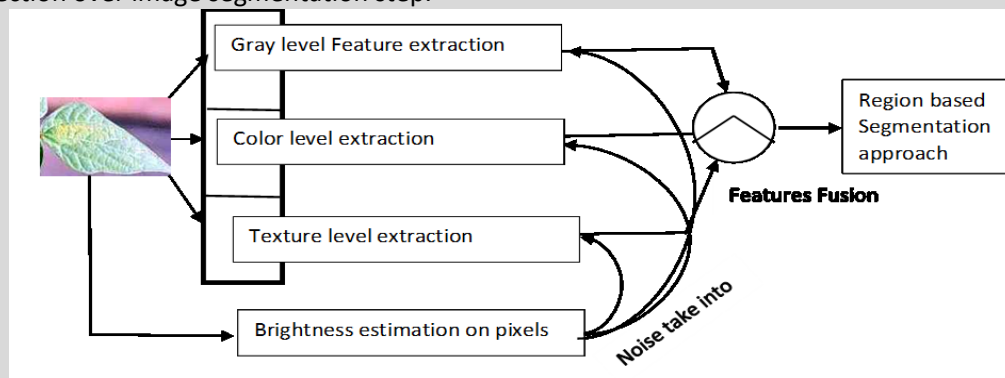


Figure 1. Modified version of a region-based segmentation approach in an evidential context

Résultats :

In order to test, examine and evaluate the performance of our segmentation techniques. Several experiments were conducted on images databases of the cowpea leaves, lettuce leaves, and cucumber disease affected leaves, with an algorithm on Visual Studio platform (language C++) with OpenCV library. Several images of various plants leaf disease in the database were collected with digital camera in the agricultural demonstration zone of Tropical Agriculture at the International Institute of Tropical Agriculture (IITA) and the experimental farms of the Faculty of Agricultural Sciences of Abomey-Calavi University in Benin. Experiments results are shown that, most of the disease area was detected, each using the three characters of gray level, color, and texture. Nevertheless, it can be observed that when noise is generated on the image during the acquisition phase, the quality of the detection is degraded, so that the disease can be more accurately detected in areas where the symptoms of the disease are not very visible yet. Figure 2, is an illustration of image segmentation result which we noted a better detection thanks to the combination of the three parameters which is a method of segmentation with extended hybrid space of image feature. The method has a great advantage compared to the methods we presented earlier because in color image, there is no change of color or change of texture and this is valid for thresholding by gray level.



Figure 2. Illustration of result of Plant Disease image segmentation on leaf

Conclusion et perspectives :

In this paper, we proposed a new segmentation method of plant disease leaf image by combining three features (gray level, color and texture) and the potential noise source using TBM Framework. Since region based approach technique segmentation method is sensitive to noise but also the starting condition. In our approach, we used the framework of transferable belief function to model the different potential sources of acquisition noise that are taken into account when extracting image features that have been negatively impacted by its noise. Hence the robustness of the proposed technique to be robust against noise. In addition, three different features of the image were extracted and merged into one by exploiting the cautious information combination rules provided by the TBM tool to enhance the sowing performance and power of the segmentation algorithm. Experimental results show that the proposed method is appropriate for dealing with plant disease leaf image segmentation and has certain robustness for noise and effective and accurate image segmentation method in the field of plant disease detection. In the future, we will further compare the potential benefits and limitations of the other categories existing method against image segmentation methods and explore a fast disease leaf image segmentation method.

Références (5 max.):

- [1] K.Thangadurai and K.Padmavathi, "Computer Vision Image Enhancement For Plant Leaves Disease Detection," IEEE, World Congress on Computing and Communication Technologies", Trichirappalli, India, pp: 173-174, 2014.
- [2] Mrunalini R. Badnakhe, Prashant R. Deshmukh, "Infected Leaf Analysis and Comparison by Otsu Threshold and k-Means Clustering" IJARCSSE, vol. 2(3), pp. 449-452, 2012.
- [3] H. Al-Hiary, S. Bani-Ahmad, M. Reyalat, M. Braik and Z. ALRahamneh, "Fast and Accurate Detection and Classification of Plant Diseases" IEEE, IJCA, vol. 17(1), pp. 31-38, March 2011.
- [4] Peng B., Zhang L., and Zhang D., "Automatic Image Segmentation by Dynamic Region Merging," IEEE Transactions on Image Processing, vol. 20, no. 12, pp. 3592-3605, 2011.
- [5] Cointault, D. Guérin, J.P. Guillemin, B.Chopinnet, 'In-Field Wheat ears Counting Using Color-Texture Image Analysis', New Zealand Journal of Crop and Horticultural Science, Vol. 36, pp.117-130, Juin 2008.
- [6] P. Smets. Decision making in the TBM : The necessity of the pignistic transformation. Int. Jour. of Approximate Reasoning, 38 :133-147, 2005.
- [7] Cointault, D. Guérin, J.P. Guillemin, B.Chopinnet, 'In-Field Wheat ears Counting Using Color-Texture Image Analysis', New Zealand Journal of Crop and Horticultural Science, Vol. 36, pp.117-130, Juin 2008.



Titre :

Segmentation-based Animal detection in aerial images

Auteurs :

Yunfei Fang^{1,2}, Shengzhi Du¹, Larbi Boubchir², Boubaker Daachi², Karim Djouani¹

¹Electrical Engineering, Faculty of Engineering and Built Environment, Tshwane University of Technology, Pretoria, South Africa

²LIASD research Lab., University of Paris 8, France

Résumé (100-200 mots) :

With the popularization of drones and accumulation of videos/images taken from aerial perspective. We investigated taking advantage of computer vision techniques to automatically process such data and benefit the field of nature conservation specifically in our research. For the purpose of detecting animals from such videos/images, we applied a bottom-up segments generating framework. Independent hierarchical figure-ground segments were automatically generated and a regression tree was trained to select the most possible object hypothesis. Despite of many challenges, such as aerial viewpoint, nature environment complexity, lack of database, the algorithm proves its efficiency in fulfilling the objective.

Mots-clefs (3 à 5 max.) :

Image segmentation, Aerial image, Min-cut, Drone.

Contexte et état de l'art

Object detection and classification has been at the center of the stage of computer vision applications. Several benchmarks are designed to evaluate the performance of different algorithms, such as MSRC, PASCAL VOC and ImageNet. The capability of computational visual reception has extended dramatically as the challenge arises. Viola and Johns in [1] reached the milestone of face detection by using computational light features (Haar) and a cascade of adboost classifiers, which makes real-time face detection possible. In [2] Pedro et al. describe the object with a collection of parts in deformable configuration. The so-called latent SVM trains the parts discriminatively but only a bounding box of the object is needed for training. Using histogram of oriented gradients (HOG), this model achieved good performance in multi-class object detection benchmarks. However, this algorithm requires a rigid model that can hardly be applied to aerial images. For example, two different models shall be trained separately in order to recognize an up-standing human and one just shows the upper-body. And the sliding-window technique, although explores in various scales, presumes a fixed aspect ratio bounding box. When observed from above, adding a third dimension, the object appearance variates so enormously that it is difficult to achieve and apply such models. Segmentation based detection algorithms may alleviate the constrains. Arbelaez et al. [3] proposed a high-quality contour detector by combining multiple local cues. Taking the advantage, Carreira and Sminchisescu [4] tried to exhaust the potential regions hierarchically by placing a set of random foreground and background seeds and formalized it to normalized cut problems. A pool of segments is generated and a regressor was trained to predict the plausibility of a segment being an object.

Travail proposé (Décrire l'objectif du travail et indiquer clairement le problème/défi technique étudié) :

This research aims at detecting terrestrial mammals from aerial videos/images taken from natural environment, specifically in typical South African landscape. Challenges are introduced by the aerial viewpoint and complexity of natural environment. As database of this kind doesn't exist yet, it is also difficult to explore existing machine learning based object detection algorithms, big amount of data is required for training and validation. From the original image, we generated a pool of hierarchical figure-ground segments by solving a set of Constrained parametric min-cuts (CPMC) problems with different seeds and unary terms [4]. A fast rejection step was taken to filter the segments. We trained a regression tree on the overlap a segment has with the hand annotated ground truth. The overlap, defined as :

$$O(S, G) = \frac{|S \cap G|}{|S \cup G|}$$

S are the segment pixels, G are the ground truth pixels, is used to illustrate the possibility of a segment being an object. Instead of regress merely on the ground truth, we took some of the high overlapped segments as training.

We experimented some of the mid-level features, trying to keep the most expressive ones, as the feature dimensions are quite limited. We used 28-dimensional region and gestalt features and 12-dimensional GLCM features. Giving a test image, such features will be extracted from each segment from the generated pool, and the highest scored prediction will be regarded as an detected object.

Résultats (Results):

From a video clip taken from a drone, with several blesbok in view with different location and posture, we annotated 85 of the mammal. We used 84 of them as training sample and 1 for testing. For each individual animal, top 100 segments overlap with the ground truth were selected as samples, which gives 7500 in total, we also experimented 84 top 1 overlap. The results gave acceptable prediction when the features are chosen effectively and with reasonable dimensions. The pictures below show the procedure and one of the prediction results.

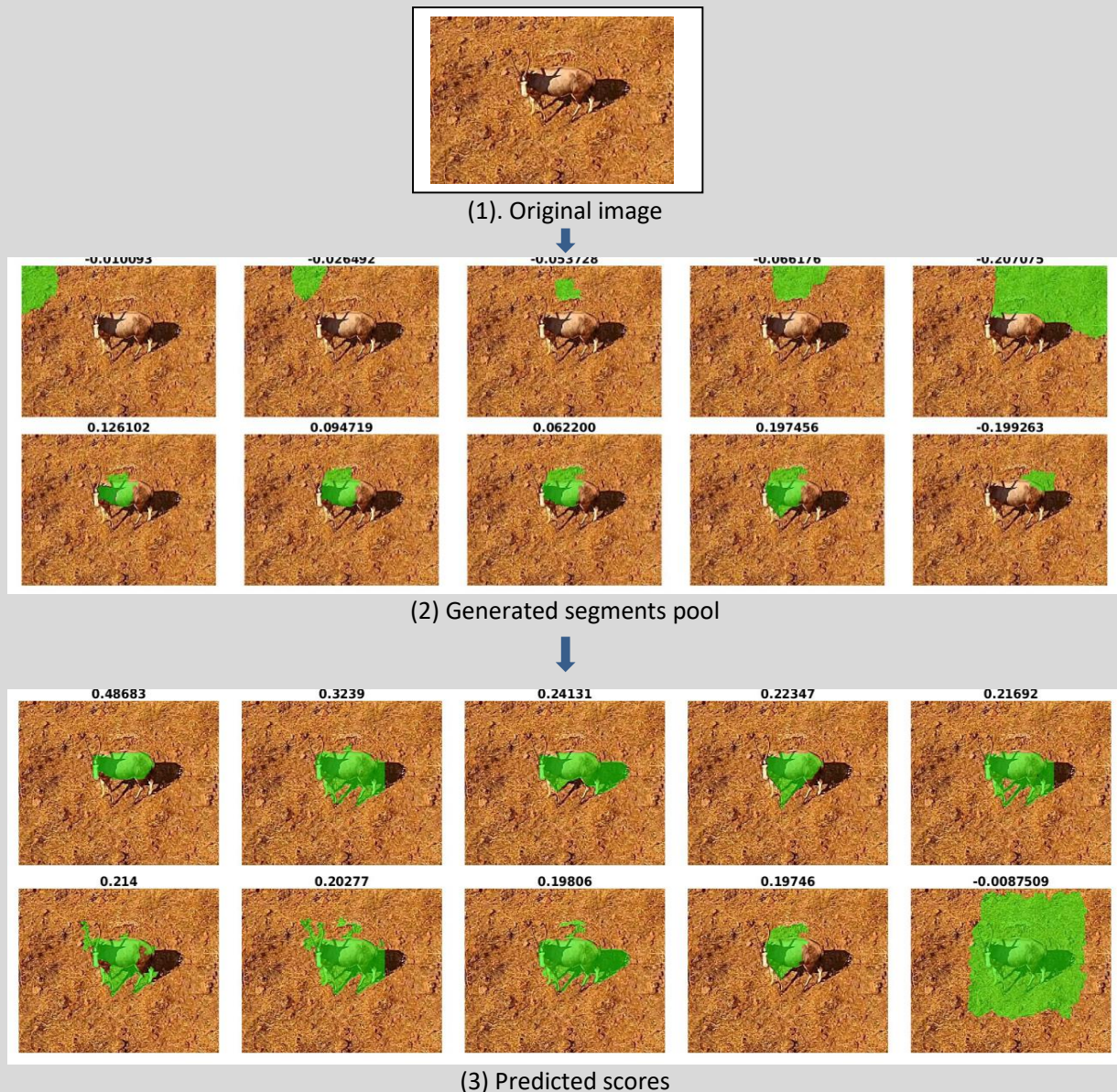


Fig.1 The algorithm procedure

Fig.1(2) shows some of the examples of the generated segments. Fig.1(3) shows the highest scored prediction and its relevant segment, the last segment from Fig.1(3) was from the 100th score, for comparison. The results verified the efficiency of the mentioned segmentation algorithm when applied to aerial images from natural environment. And the regression model works well even with limited data when provided the right features.

Conclusion et perspectives (Conclusion and perspectives :)

This approach combines hierarchical image segmentation and machine learning to detect animals in aerial images. The min-cut based segmentation algorithm generates a pool of foreground hypothesis by automatically laying the foreground and background seeds. It adapts to the challenge introduced by the natural environment, such as the homogeneous colour between the animal skin and the background. The selective features which captures mid-level properties of the object, such as convexity, continuity and texture properties are expressive, keeping the feature dimensions at a low level and avoid the disadvantage of aerial viewpoint, such as dependence in object appearance with its relative position with the camera. Even though provided with limited training data, the algorithm segments and detects the animal efficiently. Because of the ranking scheme of the algorithm, in the context of automatic object detection, it is only able to identify single object in the image. It is still left open when dealing with multiple object detection.

Références (5 max.):

- [1] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features", CVPR, 2001.
- [2] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, D. Ramanan, "Object Detection with Discriminatively Trained Part-Based Models", IEEE PAMI, vol. 32, no. 9, pp. 1627-1645, 2010.
- [3] P. Arbelaez, M. Maire, C. Fowlkes, J. Malik, "Contour Detection and Hierarchical Image Segmentation", TPAMP, vol 33, no. 5, pp. 898-916. J. Carreira and C. Sminchisescu, "Constrained Parametric min-cuts for automatic object segmentation", CVPR, 2010.
- [4] M. Maire, P. Arbelaez, C. Fowlkes and J. Malik, "Using contours to detect and localize junctions in natural images", CVPR, 2008.
- [5] M. Maire, P. Arbelaez, C. Fowlkes and J. Malik, "Using contours to detect and localize junctions in natural images", CVPR, 2008.



Titre :

Détection des attaques de présentation en reconnaissance faciale en combinant l'apprentissage profond et les motifs binaires locaux

Auteurs :

Tarek Ben Charrada, Wael Elloumi, Aladine Chetouani

Résumé (100-200 mots) :

Face recognition has been widely applied as a user authentication method due to its nonintrusive behaviour. However, spoofing those systems can be a relatively easy task, as it only requires a picture, a video-replay, or a 3D mask to bypass such a security layer and illegally earn access to what may be confidential information. Hence, another security layer is mandatory. Face anti-spoofing module is the extra security layer used to prevent such an illegal access attempt. In this research, we demonstrate the ability of our multi-input architecture that combines a pre-trained CNN model based on transfer learning and the Local-binary patterns (LBP) descriptor to make accurate near real-time predictions on mobile devices. Experimental results reveal that our approach achieves a perfect discrimination on high quality videos and a HTER of 2.63% when classifying low quality videos.

Mots-clefs (3 à 5 max.) :

Face recognition, Deep learning, Transfer learning, Local Binary Patterns, Liveness detection.

Contexte et état de l'art :

Facial recognition is a biometric modality that has been recently gaining interest for both identification and authentication purposes. Thanks to its convenience in today's digital age, this technology is more and more integrated into the daily lives of consumers. Moreover, since it only requires a front face camera, it has paved the way for selfie-based authentication on smartphones, leading to its adoption in many critical applications such as mobile banking, payment apps and border control. However, the overall security of face recognition-based systems represents the major concern of users. Indeed, the detection of spoofing attacks, also called direct attacks or presentation attacks, remains a challenging task that has gained attention of the research community. Biometric anti-spoofing methods can be divided into 3 groups: dynamic, static and multimodal countermeasures. Dynamic anti-spoofing methods rely on motion analysis over a face video sequence to detect physiological signs of liveness like eye blinking, eye movement or lip movement. Motion cues can also be based on the subtle relative movements between facial parts, the overall motion correlation between the face and the background regions, or the motion estimation to detect planar media-produced attacks such as prints or screens. However, involving the collaboration of the user is not very adequate to the sought non-intrusive characteristic of biometric authentication systems. While dynamic approaches rely on the video data, static methods focus on a single instance of the face. In general, these techniques are faster and less intrusive. The key assumption behind their intuition is that a fake face is likely to have lower image quality than a genuine one. Therefore, static approaches rely on face appearance analysis, i.e., texture analysis-based techniques, using texture information or Image Quality Measures (IQMs). Because of the lower accuracy of static methods, some works have combined several image quality measures (IQMs) to detect spoofing attacks [1]. Multimodal techniques have also been explored as a countermeasure to spoofing attacks. They rely on several biometric modalities for a stronger authentication, such as the combination of face and voice, the correlation between voice and lips motions or the combination of face, voice and iris, etc. However, such solutions may entail challenges like inconvenience, cost or harm of user experience. In the present work, we present a static approach that combines a pre-trained CNN model based on transfer learning and the Local-binary patterns (LBP) descriptor.

Travail proposé (Décrire l'objectif du travail et indiquer clairement le problème/défi technique étudié) :

Our purpose is to build an anti-spoofing solution for face recognition authentication on mobile devices. Thus the two major criteria of our application are classification performance and computational cost.

Architecture

The proposed architecture process two aspects of inputs simultaneously: the raw face images and their respective Local-binary patterns (LBP) descriptor. Face images pass through a set of convolutional layers which we deploy as feature extractors. We use the same architecture as the VGG-16 [2] to retrieve a deep representation from raw faces. After these convolutional layers, stands a 265 fully connected layer (FC). The output of the convolutional layers is concatenated with the LBP histograms of the frames before being passed to the FC layer as shown in figures 1. For the decision making process we deploy a Linear Discriminant Analysis (LDA) classifier.

Training

For the feature extractor training process we use the following architecture.

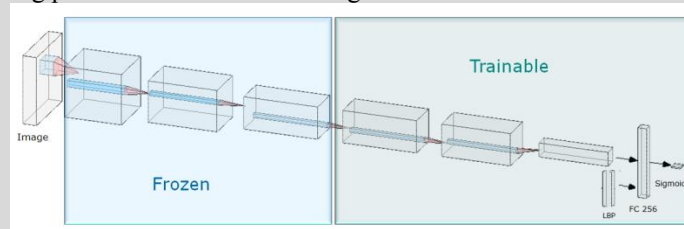


Figure 1: Training procedure

First, we froze the first 3 blocks of the VGG and we trained the remaining layers. As the VGG-16 model was already trained on the Imagenet dataset, we proceeded by fine tuning its layers using the training subset of our used datasets. Then, we joined the output of our pre-trained CNNs with the LBP histograms and trained the LDA classifier. We used Singular value decomposition as a solver and a rank estimation threshold of 10^{-4} .

Résultats :

Datasets

We evaluated our method on two publicly available databases:

Replay Attack [3]: It incorporates 1300 videos of 50 identities. In addition to real-access samples, it includes three types of attack: photos printed on paper, videos displayed on an iPhone 3GS screen, and those displayed on a high resolution screen using an iPad.

Replay Mobile [4]: It consists of 1190 videos of 40 subjects. Likewise, Replay Mobile is divided into subsets using the same structure as Replay Attack, and it also includes different lighting conditions. The videos were recorded using an iPad Mini2 and a LG-G4 smartphone, with a resolution of (1280 x 720) for at least 10 seconds.

Results

Tables 1 and 2 shows the obtained results on Replay Attack and Replay Mobile respectively. These tables report video based results. We used majority vote to classify each video. Our proposed approach reports a perfect classification on Replay Mobile database. On the Replay Attack database, The FASNet model [5] reports a Half total error rate (HTER) of 1.2% which is better than the HTER reached by our approach (2.63%). However, it is worth noting that our method denies all the imposter attacks.

	Predicted real	Predicted attack
Real	76	4
Attack	0	400

Table 1: Replay Attack confusion matrix

	Predicted real	Predicted attack
Real	110	0
Attack	0	192

Table 2: Replay Mobile confusion matrix

Conclusion et perspectives :

In this paper, we present a new approach that leverages transfer learning in convolutional neural networks and Local binary patterns (LBP) to address face anti-spoofing on mobile devices. Our experimental results showed a perfect discrimination on high quality videos and a HTER of 2.63% when classifying low quality videos. For future work, we intend to test our approach on other databases commonly used in the literature for anti-spoofing purposes, in order to study the generalization of our approach

Références (5 max.):

- [1] Fourati, E., Elloumi, W., and Chetouani, A. (2017). Face anti-spoofing with image quality assessment. In 2017 2nd International Conference on Bio-engineering for Smart Technologies (BioSMART), pages 1-4.
- [2] Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. CoRR, abs/1409.1556.
- [3] Chingovska, I., Anjos, A., and Marcel, S. (2012). On the effectiveness of local binary patterns in face anti-spoofing.
- [4] Costa-Pazo, A., Bhattacharjee, S., Vazquez-Fernandez, E., and Marcel, S. (2016). The replay-mobile face presentation-attack database. In Proceedings of the International Conference on Biometrics Special Interests Group (BioSIG)
- [5] Lucena, O., Junior, A., Hugo G Moia, V., Souza, R., Valle, E., and De Alencar Lotufo, R. (2017). Transfer learning using convolutional neural networks for face anti-spoofing. pages 27-34.



Titre :

Amélioration de la robustesse de la détection d'images altérées pour la compression

Auteurs :

Boubacar Diallo, Laboratoire XLIM, boubacar.diallo@univ-poitiers.fr

Thierry Urruty, Laboratoire XLIM, thierry.urruty@univ-poitiers.fr

Pascal Bourdon, Laboratoire XLIM, pascal.bourdon@univ-poitiers.fr

Christine Fernandez-Maloigne, Laboratoire XLIM, christine.fernandez@univ-poitiers.fr

Résumé (100-200 mots) :

La vérification de l'origine et de l'authenticité des images numériques impose de nombreuses contraintes aux algorithmes de détection de manipulations malveillantes par rapport aux traitements usuels. La compression étant le type de post-traitement le plus couramment utilisé, nous proposons dans notre travail un cadre robuste contre cette transformation particulière. Notre démarche repose sur 2 étapes : l'identification du modèle d'appareil photo utilisé pour l'acquisition de l'image, à l'aide de techniques d'apprentissage profond ; puis la détection des falsifications éventuelles, en prenant en compte les facteurs de compression possible. Nos expériences sur des ensembles de données de référence montrent la contribution de notre proposition à la fois à l'identification de modèles d'appareils photos et à la détection de falsifications d'images par rapport aux approches récentes de la littérature.

Mots-clefs (3 à 5 max.) :

Détection d'images altérées
Identification de modèle d'appareil photo
Compression avec perte
Réseaux de neurones convolutifs

Contexte et état de l'art :

Nous présentons dans cet article un cadre global qui fournit une solution robuste pour l'identification d'un modèle d'appareil photo ayant servi à une acquisition et pour la détection d'images altérées. La motivation de ce travail vient du fait que la plupart des images sont acquises sur Internet et que, parmi celles-ci, une proportion importante subit des transformations telles que la compression, le bruitage, le redimensionnement et/ou le filtrage, appliquées à la fois avant et après une éventuelle modification. Il est donc essentiel que les algorithmes de détection d'images altérées, permettant de valider l'intégrité d'une image, soient d'une grande robustesse face à ces différentes manipulations. La compression étant le type de transformation le plus couramment appliqué lorsque des personnes partagent des images sur Internet, nos expériences portent sur cette transformation.

Travail proposé (Décrire l'objectif du travail et indiquer clairement le problème/défi technique étudié) :

Ce travail est divisé en deux parties comme indiqué dans la figure 1. Dans la première partie, nous détaillons notre approche d'apprentissage en profondeur pour identifier les modèles d'appareils photos. La seconde partie explique comment cette approche est incluse dans un cadre global de détection d'images frauduleuses. L'objectif étant de détecter toutes modifications d'images partagées sur Internet, nous proposons une technique globale d'augmentation des données d'apprentissage permettant au modèle CNN (réseaux de neurones convolutifs) d'exploiter des images soumises aux mêmes transformations que celles utilisées couramment par les utilisateurs, telles que la compression avec perte, le bruitage, le redimensionnement et / ou le filtrage. Les expériences montrent comment notre schéma permet d'obtenir plus de robustesse à la compression.

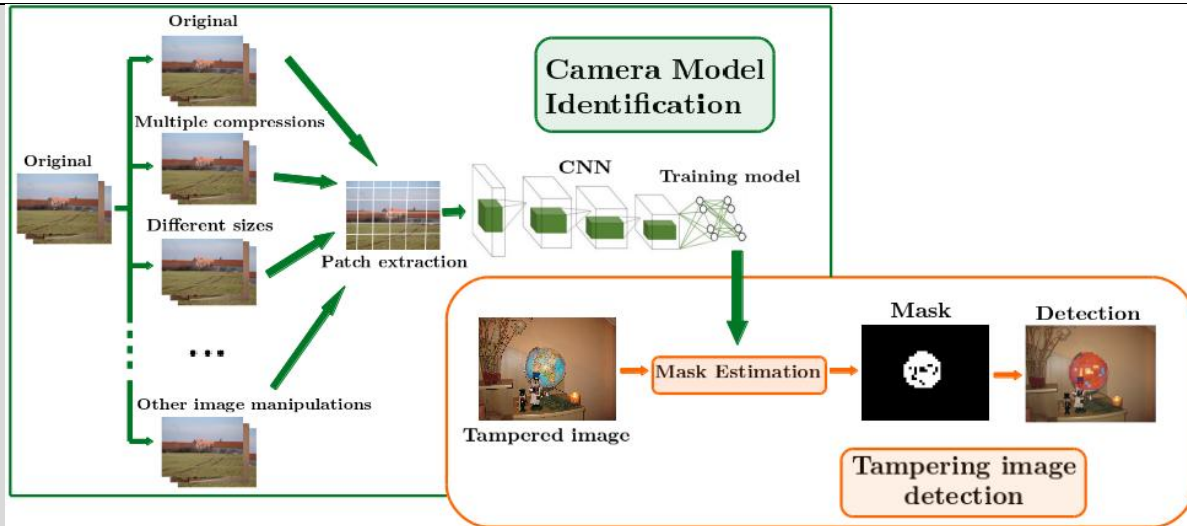
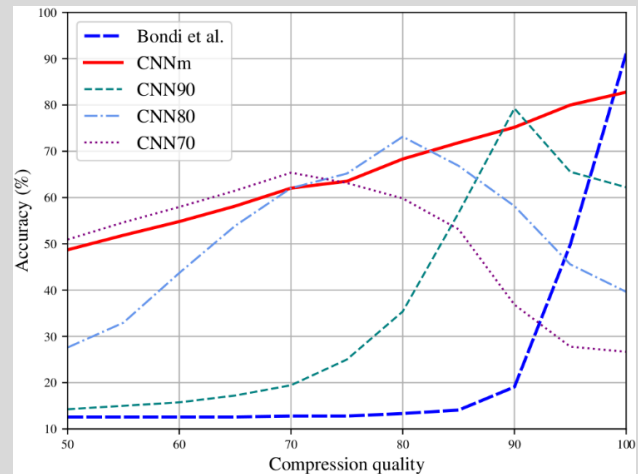


Figure 1 : Architecture de notre framework comprenant la phase d'apprentissage d'identification de modèles d'appareils photos et le procédé de détection d'images frauduleuses.

Résultats :

Ces travaux de recherche sont une étude préliminaire mettant en évidence l'importance de la compression en tant que technique de manipulation d'images dans le cadre de la mesure d'intégrité d'une image. Pour évaluer l'influence de la compression, toutes les images des jeux de données choisis pour l'étape de reconnaissance d'un type d'appareil photo ayant servi à l'acquisition d'une image sont compressées avec différents facteurs de qualité (FQ) : **90%**, **80%** et **70%**. Les CNN formés avec ces FQ sont nommés respectivement **CNN90**, **CNN80**, **CNN70** et **CNNm** pour les données compressées mixtes.

La figure 2 montre l'influence de la compression sur la précision d'identification d'un modèle. Ce résultat montre que le CNN formé uniquement sur l'image "originale" pour l'identification du modèle d'appareil photo [1] n'est pas robuste à la compression. Elle confirme le fait qu'un modèle formé par un CNN avec un facteur de qualité de compression spécifique ne donne une précision supérieure que pour ce facteur de qualité. Notre approche (**CNNm**) donne en moyenne des résultats performants lorsqu'on utilise un ensemble d'images compressées à différents facteurs de qualité.



Conclusion et perspectives :

Dans cet article, nous proposons une approche robuste basée sur l'apprentissage profond pour l'identification du modèle d'appareil photo ayant servi à l'acquisition d'une image et la détection d'images falsifiées. Cette approche peut s'étendre à d'autres manipulations couramment utilisées par les utilisateurs partageant des images. Nous avons testé notre méthode sur la compression à différents facteurs de qualité et avons montré que notre approche surpasse globalement les approches existantes dans la littérature [1, 2]. Dans nos travaux futurs, nous comptons augmenter encore la robustesse de notre approche en étudiant les nœuds d'activation de réseaux de neurones afin de mieux comprendre les artefacts permettant d'identifier les modèles d'appareil photo.

Références (5 max.) :

- [1] BONDI, Luca, et al. First steps toward camera model identification with convolutional neural networks. *IEEE Signal Processing Letters*, 2017, 24.3: 259-263.
- [2] BONDI, Luca, et al. Tampering Detection and Localization Through Clustering of Camera-Based CNN Features. In: *CVPR Workshops*. 2017. p. 1855-1864.



Titre :

Etude de l'apport de la reconstruction des régions occultées du visage pour la reconnaissance des expressions

Auteurs :

Delphine Poux [1,2], Benjamin Allaert [1], Nacim Ihaddadene [2], Ioan Marius Bilasco [1], Chaabane Djeraba [1]

[1] Centre de Recherche en Informatique Signal et Automatique de Lille, Univ. Lille, CNRS, Centrale Lille, UMR 9189 - CRISTAL -, F-59000 Lille, France

[2] ISEN Lille, Yncrea Hauts-de-France, France

Résumé (100-200 mots) :

La reconnaissance des expressions faciales en contexte d'interaction non contrôlée est, encore aujourd'hui, confrontée à un certain nombre de défis qui réduisent considérablement les performances des systèmes. Parmi ces défis, les occultations induites par les variations de pose de la tête ou par des objets externes (accessoires, mains, ...) impliquent une perte d'informations sur le visage. Différentes méthodes proposées dans la littérature tendent à résoudre ces problèmes soit en reconstruisant les régions occultées, soit en exploitant uniquement les informations au sein des régions visibles du visage. Dans cet article, nous étudions l'apport de la reconstruction des régions occultées, par rapport à l'exploitation efficace de l'information contenue dans les régions visibles du visage lors de la reconnaissance des expressions faciales.

Mots-clefs (3 à 5 max.) :

Expressions faciales, Occultations, Mouvement, Reconstruction

Contexte et état de l'art :

Les systèmes de reconnaissance des expressions faciales proposés dans la littérature obtiennent de très bonnes performances dans des conditions contrôlées (pose fixe, visage frontal). Cependant, dans un contexte d'interaction naturelle, où la personne est libre de ses mouvements, les variations de pose ou les gestes occultant le visage engendrent une perte d'information. Pour pallier cette perte d'information, deux familles de méthodes se distinguent : une basée sur la reconstruction de l'information manquante et une basée sur l'exploitation optimale des informations visibles, non-impactées par l'occultation. Les méthodes basées sur la reconstruction amènent le visage dans un contexte idéal d'analyse. Cependant, bien que ces méthodes soient efficaces pour la reconnaissance faciale, elles restent encore mal adaptées pour la reconnaissance des expressions faciales, car elles ne préservent pas efficacement la géométrie du visage [1]. Quant aux méthodes exploitant uniquement les régions visibles, elles garantissent que les données extraites sur le visage ne subissent aucun biais induit par un pré-traitement.

Travail proposé (Décrire l'objectif du travail et indiquer clairement le problème/défi technique étudié) :

Dans cette étude nous comparons l'apport d'une méthode de reconstruction du mouvement par rapport à une méthode de sélection du mouvement dans le cadre de la reconnaissance des expressions en présences d'occultations. Nous exploitons le mouvement facial afin de reconnaître les expressions en nous appuyant sur le descripteur LMP et le modèle de segmentation facial utilisant 25 régions, introduits en [3]. Poux et al. [2] montrent qu'un descripteur de mouvement est pertinent en présence d'occultations du visage car le mouvement extrait à proximité des régions occultées contient des informations importantes induites par la propagation du mouvement.

La méthode de reconstruction proposée estime le mouvement des régions occultées en s'appuyant sur le mouvement de la région symétrique associée. La méthode de sélection suit le processus de sélection de régions significatives introduit dans [2]. Cette méthode consiste à analyser l'importance de chaque région du visage pour une expression et une occultation spécifique. Ceci implique de construire un modèle d'apprentissage par expression et par occultation. Afin de suivre un protocole de validation commun, un modèle d'apprentissage par expression est construit pour les deux méthodes. En nous inspirant de [2], nous mettons en place un protocole expérimental permettant de comparer la méthode de reconstruction et la méthode de sélection. La base de données CK+ [4] est utilisée pour comparer ces deux méthodes. Le jeu de données utilisé contient 374 séquences provenant de 123 participants. Les séquences couvrent six

expressions (colère, dégoût, peur, joie, tristesse et surprise). Les visages sont enregistrés dans de parfaites conditions (pose fixe, visage frontal). D'importantes occultations synthétiques, représentées dans la Figure 1, sont ensuite appliquées sur les visages afin de mieux comparer les performances des deux méthodes d'analyse.

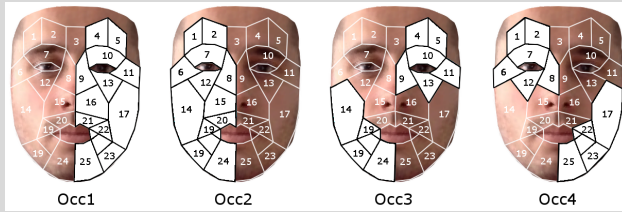


Fig.1 – Représentation des occultations appliquées sur les données de CK+ [Occ 1 et 2: variations de pose ; Occ 3 et 4: objets externes (lunettes, cheveux, main)].

Résultats :

Les performances obtenues pour l'ensemble des occultations sont représentées dans le tableau 1. Le taux de reconnaissance correspond à la moyenne des taux obtenus sur l'ensemble des occultations par expression. Les taux de reconnaissances sont obtenus en utilisant un classifieur SVM et un protocole de validation croisée en 10-fold. Concernant la méthode par reconstruction, le modèle d'apprentissage est construit en s'appuyant sur les informations des 25 régions. Quant à la méthode exploitant les régions visibles, pour chaque expression, seules les 5 meilleures régions pour chaque expression sont utilisées. Ces 5 régions sont définies par [2], où les auteurs montrent qu'en exploitant uniquement 5 régions par expression, les performances obtenues sont similaires aux performances obtenues en exploitant l'ensemble des 25 régions.

Méthodes	Joie	Peur	Surprise	Colère	Dégoût	Tristesse	Moyenne
Reconstruction (25R)	94.25% Std : 2.6	77.46% Std : 6.3	96.13% Std : 1.0	84.81% Std : 2.8	91.75% Std : 2.6	86.78% Std : 3.8	88.53%
Régions visibles (5R)	94.92% Std : 0.1	89.48% Std : 1.0	96.75% Std : 1.0	86.40% Std : 2.0	95.05% Std : 2.0	90.89% Std : 1.8	92.25%

Les résultats montrent qu'en s'appuyant sur un sous-ensemble de 5 régions du visage pour chaque expression, la méthode exploitant uniquement les régions visibles obtient de meilleures performances que celle basée sur la reconstruction. Cependant, celle-ci nécessite de créer un modèle d'apprentissage par expression et par occultation. Quant à la méthode de reconstruction, basée sur les 25 régions du visage, elle a l'avantage de nécessiter un unique modèle d'apprentissage. Il est toutefois important de considérer que les performances de ces méthodes dépendent fortement de leur capacité à reconstruire les régions occultées du visage tout en conservant fidèlement la géométrie faciale.

Conclusion et perspectives :

Dans le cadre de la reconnaissance des expressions faciales, nous avons comparé deux méthodes, une par reconstruction et une par sélection, visant à réduire l'impact de la perte d'information induite par des occultations. Dans ce papier, nous montrons que pour analyser les expressions faciales en présence d'occultations partielles du visage, il est préférable d'exploiter uniquement l'information contenue dans les régions visibles au lieu de reconstruire les régions occultées. Cette solution permet de s'affranchir des problèmes d'estimation des méthodes de reconstruction, qui peuvent avoir de mauvaises répercussions sur la reconnaissance des expressions. De plus, elle permet d'obtenir des performances significatives tout en s'appuyant sur un sous-ensemble de régions, ce qui met en évidence la robustesse de cette méthode en présence d'importantes occultations. Cependant, la construction d'un modèle d'apprentissage par expression et par occultation est nécessaire, contrairement à la méthode basée sur la reconstruction qui nécessite un seul modèle contenant toutes les régions.

Comme perspective, dans un premier temps, nous souhaitons étudier de manière plus fine l'apport de la reconstruction du mouvement au niveau de chaque région. Cela permettra d'avoir une meilleure compréhension du gain apporté par la reconstruction sur les différentes régions du visage en fonction des différentes expressions et des occultations subies. Dans un second temps, nous allons approfondir notre étude en présence d'occultations induites par de réelles variations de pose et d'objets externes.

Références (5 max.):

- [1] Allaert, B., Mennesson, J., Bilasco, I. M. & Djeraba, C. (2018). Impact of the face registration techniques on facial expressions recognition. *Signal Processing: Image Communication*, 61, 44-53.
- [2] Poux, D., Allaert, B., Mennesson, J., Ihaddadene, N., Bilasco, I. M., & Djeraba, C. (2018). Mastering Occlusions by Using Intelligent Facial Frameworks Based on the Propagation of Movement. In *International Conference on Content-Based Multimedia Indexing (CBMI)*.
- [3] Allaert, B., Mennesson, J., Bilasco, I. M. & Djeraba, C. (2018). Advanced local motion patterns for macro and micro facial expression recognition. *arXiv preprint arXiv :1805.01951*.
- [4] Lucey, P., Cohn, J. F., Kanade, T., Saragih, J., Ambadar, Z., & Matthews, I. (2010). The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression. In *Computer Vision and Pattern Recognition Workshops (CVPRW), Computer Society Conference on* (pp. 94-101). IEEE.



Titre :

Protection de la vie privée par partage hybride de photos sur les réseaux sociaux

Auteurs :

Sébastien BEUGNON, Pauline Puteaux et William Puech
LIRMM, Univ. Montpellier, CNRS, Montpellier, France
{sebastien.beugnon, pauline.puteaux, william.puech}@lirmm.fr

Résumé (100-200 mots) :

Le développement des réseaux sociaux soulève de nombreuses questions relatives à la protection des images. En particulier, des conflits de protection de la vie privée multipartite peuvent avoir lieu lors de la publication d'une photo par son propriétaire. En effet, les paramètres de confidentialité appliqués à cette image sont ceux de son propriétaire et les personnes y figurant n'interviennent pas dans le processus.

Dans ce papier, nous proposons une méthode de partage hybride de photos sur les réseaux sociaux pour répondre à ce problème. Basée sur l'approche de partage de secret de Shamir appliquée aux images, cette solution permet d'assurer le respect de la vie privée des utilisateurs, comme l'attestent les résultats expérimentaux obtenus.

Mots-clefs (3 à 5 max.) :

Sécurité multimédia, partage d'image secrète, protection de la vie privée, réseaux sociaux.

Contexte et état de l'art :

Ces dernières années, la sécurité des données visuelles est devenue un sérieux problème. Avec le développement d'Internet, de plus en plus de données visuelles – images et vidéos – sont transmises sur les réseaux et stockées sur le *cloud*. Les réseaux sociaux sont particulièrement populaires avec plus de deux milliards d'utilisateurs actifs à travers le monde. Les données visuelles transitant sur ces réseaux sont généralement à caractère personnel. Générées en large quantité et par les utilisateurs eux-mêmes, leur sécurité est constamment menacée. En particulier, le respect de la vie privée de chacun lors du partage de données impliquant plusieurs utilisateurs constitue un problème majeur, qui peut être formulé comme suit. Alice prend une photo avec son groupe d'amis et la publie sur sa page personnelle. L'ensemble de son réseau a alors accès à cette photo suivant les paramètres de confidentialité qu'elle a elle-même définis. Par ailleurs, ses amis, présents sur la photo, n'interviennent pas dans la procédure de publication. En effet, ils ne sont pas consultés et n'ont pas à donner leur accord avant la publication de la-dite photo, contenant pourtant des informations personnelles à leur sujet [1]. Dans ce contexte, il est nécessaire de proposer une solution efficace afin de gérer ces conflits de protection de la vie privée multipartite.

Les méthodes de partage d'image secrète (*secret image sharing*) peuvent être utilisées pour répondre à cette problématique. Inspirées des méthodes de partage de secret développées indépendamment par Blakley [2] et Shamir [3] en 1979, elles permettent de partager une image entre n utilisateurs de façon sécurisée [4]. Chaque utilisateur reçoit une information issue de ce partage sous la forme d'une image appelée *share*. Cette *share* est personnelle, unique et semble visuellement avoir été générée aléatoirement. L'image originale ne peut alors être reconstruite qu'après réunion d'au moins k de ces *shares* avec $k \leq n$. Avec $k - 1$ *shares*, il est en effet impossible d'obtenir quelque information sur le contenu original de l'image. Le paramètre k peut être plus ou moins élevé, suivant le niveau de confiance au sein du groupe de partage. Le processus de partage est illustré en figure 1.

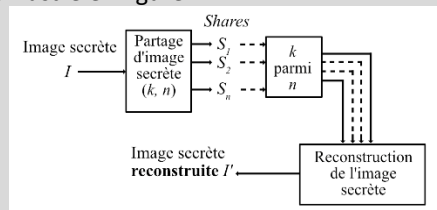


Figure 1 - Schéma de partage d'image secrète (*secret image sharing*).

Dans le contexte qui nous intéresse dans ce papier, le partage d'une photo représentant plusieurs personnes sur les réseaux sociaux peut alors s'effectuer comme suit. Alice et ses amis ont défini au préalable un niveau de confiance au sein de leur groupe. Via le réseau social, ils ont décidé d'autoriser la publication d'une photo les représentant si au moins k personnes parmi les n présentes donnent leur accord. Lorsqu'Alice souhaite publier une photo avec son groupe d'amis sur sa page personnelle, ses amis sont interrogés pour savoir s'ils acceptent sa divulgation. Si k d'entre eux sont d'accord,

alors la photo est publiée complètement en clair, et dans le cas contraire, le respect de leur vie privée est maintenu en rendant non visible leur visage.

Travail proposé (*Décrire l'objectif du travail et indiquer clairement le problème/défi technique étudié*) :

À travers ce papier, nous proposons une méthode de partage hybride de photos sur les réseaux sociaux assurant le respect de la vie privée des utilisateurs en protégeant visuellement leur visage. Dans le scénario proposé, n utilisateurs au sein d'un groupe commencent par définir un niveau de confiance.

Cela se traduit par le choix commun d'un paramètre k , tel que $1 < k \leq n$, indiquant le nombre minimum d'utilisateurs devant donner leur accord pour la publication de toute image les représentant. Le schéma général de la phase d'encodage de notre méthode est illustré en figure 2.

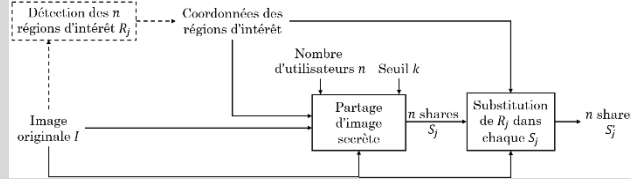


Figure 2 - Schéma général de la phase d'encodage.

L'image originale prise en entrée correspond à une photo représentant n utilisateurs. À l'aide d'un algorithme de détection des visages, des régions d'intérêt R_j , avec $j \in \llbracket 1; n \rrbracket$ associées à chaque utilisateur sont identifiées dans la photo. Les coordonnées de ces régions sont alors utilisées pour délimiter les zones de l'image à protéger par le partage d'image secrète. Le reste de l'image est public.

Pour protéger ces n régions d'intérêt, nous proposons d'appliquer la méthode de partage de secret de Shamir [3], pixel par pixel. Cette méthode utilise des polynômes définis sur un corps fini F_q tel que $|F_q| = q$, avec q un nombre premier respectant $0 < k \leq n < q$. Un polynôme de degré $(k - 1)$ est construit tel que $(k - 1)$ entiers positifs sont aléatoirement choisis afin de former l'ensemble $A = \{a_i\}$ avec $i \in \llbracket 1; k - 1 \rrbracket$, $a_i < q$ et a_0 correspondant à la valeur de chaque pixel à partager secrètement :

$$f(x) = \sum_{i=0}^{k-1} a_i \times x^i \text{ mod } q. \quad (1)$$

Un identifiant unique (ID) x_j , avec $0 < x_j < q$ et $j \in \llbracket 1; n \rrbracket$, est alors attribué à chaque utilisateur. La valeur secrète associée à chaque pixel correspond alors à $f(x_j)$. À la fin du processus, chacun des n utilisateurs reçoit un ID ainsi qu'une *share* S_j de la même taille que l'image originale. S_j est constituée de l'ensemble des pixels partagés, non visibles en clair, des n régions d'intérêt, et des pixels de l'image originale de la région publique. Ensuite, la région d'intérêt R_j en clair associée à l'utilisateur d'ID x_j se substitue à la région correspondante dans la *share* S_j pour obtenir la *share* finale S'_j .

Lorsque le nombre d'utilisateurs participant au décodage k' est inférieur au seuil k , alors la méthode utilise les k' *shares* $\{S'_j\}$ où $j \in \llbracket 1; n \rrbracket$, pour substituer dans l'image les régions protégées par les k_0 régions d'intérêt R_j en clair associées à ces k' participants et ainsi générer l'image I' .

Les régions d'intérêt des utilisateurs ne participant pas au décodage restent protégées. Dès lors que le nombre d'utilisateurs k' dépasse le seuil k , la méthode utilise une interpolation de Lagrange pour reconstruire les pixels des $n - k'$ régions d'intérêt associées aux non participants. Par ailleurs, les régions d'intérêt associées aux k_0 utilisateurs participant à la reconstruction sont en clair dans leur propre *share*. Elles se substituent dans l'image reconstruite pour former l'image I' . Cette image est alors visuellement identique à l'image originale ($I' = I$). En effet, suffisamment d'utilisateurs ont donné leur accord pour la publication de la photo originale (selon le niveau de confiance défini au préalable au sein du groupe). Le schéma général de la phase de décodage est illustré en figure 3.

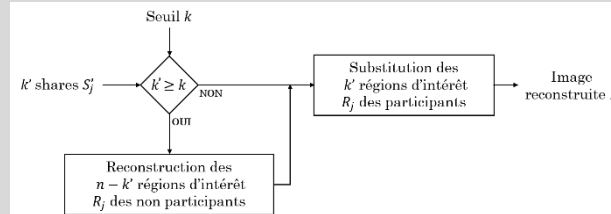


Figure 3 – Schéma general de la phase de décodage.

Résultats :

La figure 4 présente un exemple d'application de la méthode proposée avec les paramètres $(k, n) = (5, 8)$, où $n = 8$ utilisateurs ont été identifiés après la détection des régions d'intérêt. La reconstruction est alors possible lorsqu'au moins $k = 5$ utilisateurs sont d'accord. La figure 4.a illustre les régions d'intérêt détectées à l'aide d'un algorithme de détection des visages [5]. Dans l'approche proposée, une *share* est publiée sur le réseau social et les utilisateurs reçoivent chacun une *share* dont les régions d'intérêt sont protégées, à l'exception de la région d'intérêt associée à cet utilisateur. La figure 3.b représente la *share* publique tandis que la figure 4.c illustre la *share* associée à l'utilisateur d'ID $= x_2$. Tant que le nombre

d'utilisateurs participant à la reconstruction k' reste inférieur au seuil k fixé lors du partage, par exemple $k' = 3$ en figure 4.d, alors l'image reconstruite contient uniquement le contenu public et les régions d'intérêt des utilisateurs (IDs $= \{x_1, x_2, x_4\}$) ayant accepté la visualisation du contenu de l'image. Les régions d'intérêt associées aux utilisateurs qui ne participent pas à la reconstruction restent donc protégées intégralement. Enfin, lorsqu'au moins $k' \geq k$ utilisateurs se réunissent, par exemple $k' = 5$ en figure 4.e, alors le nombre d'utilisateurs devient suffisant pour que le contenu intégral de l'image originale soit révélé. Les k' régions d'intérêt associées aux utilisateurs participants (IDs $= \{x_1, x_3, x_5, x_7, x_8\}$) se substituent en clair aux régions protégées correspondantes dans l'image et les $n - k'$ régions d'intérêt restantes (IDs $= \{x_2, x_4, x_6\}$), sont reconstruites à l'aide de la méthode de partage d'image secrète.



Figure 4 – Illustration de la méthode proposée avec les paramètres (5, 8) : a) Image originale après détection des régions d'intérêt (visages, en vert), b) Share publiée sur les réseaux sociaux, c) Share associée à l'utilisateur d'ID = x_2 , d) Image originale partiellement reconstruite, après réunion de $k' = 3$ utilisateurs (ici, IDs $\{x_1, x_2, x_4\}$), e) Image originale reconstruite par partage hybride, après réunion d'au moins $k' = 5$ utilisateurs (ici, IDs $= \{x_1, x_3, x_5, x_7, x_8\}$).

Conclusion et perspectives :

Dans ce papier, nous avons proposé une nouvelle méthode de partage hybride de photos sur les réseaux sociaux, assurant le respect de la vie privée des utilisateurs. Les n utilisateurs choisissent d'un commun accord de permettre la visualisation en clair de toutes les régions d'intérêt lorsqu'au moins k d'entre eux ont autorisé la reconstruction. Si ce seuil n'est pas franchi, alors seulement les régions d'intérêt rattachées aux utilisateurs participant à la reconstruction sont révélées. Les résultats expérimentaux présentés valident notre approche. De plus, notre méthode peut étendre ses usages à un cas plus classique de protection de la vie privée pluripartite où tous les utilisateurs doivent donner leur consentement pour autoriser la reconstruction. Pour cela, il suffit d'employer un seuil $k = n$ lors du partage de l'image secrète.

Dans de futurs travaux, nous sommes intéressés par l'utilisation du partage d'image secrète hiérarchique appliqué à la gestion des conflits de protection de la vie privée pluripartite. Le principe général peut alors être décrit comme suit. On suppose que n utilisateurs sont présents sur une photo. Ces n utilisateurs sont divisés en m groupes de n_i personnes, avec $1 \leq i \leq m$. Chacun de ces groupes définit un niveau de confiance k_i qui lui est propre, et participe à un protocole de partage d'image secrète (k_i, n_i) , pour les régions d'intérêt correspondant à leur groupe. Par ailleurs, les n utilisateurs sont aussi impliqués dans le partage de la photo dans son intégralité. Il est alors nécessaire de demander que k utilisateurs dans chaque groupe donnent leur accord pour permettre la visualisation de toute la photo.

Références (5 max.):

- [1] J. M. Such et N. Criado. Multiparty privacy in social media. *Communications of the ACM*, 61(8) :74-81, 2018.
- [2] G. R. Blakley. Safeguarding cryptographic keys. Dans *Proceedings of the national computer conference*, volume 48, pages 313-317, 1979.
- [3] A. Shamir. How to share a secret. *Communications of the ACM*, 22(11) :612-613, 1979.
- [4] C.-C. Thien et J.-C. Lin. Secret image sharing. *Computers & Graphics*, 26(5) :765-770, 2002.
- [5] R. Lienhart et J. Maydt. An extended set of Haar-like features for rapid object detection. Dans *IEEE International Conference on Image Processing (ICIP)*, pages 900-903. IEEE, 2002.



Titre :

Etude de l'influence du contenu vidéo dans une transmission Pseudo-Analogique de type SoftCast

Auteurs :

Anthony TRIOUX, François-Xavier COUDOUX, Patrick CORLAY, Mohamed GHARBI

Résumé (100-200 mots) :

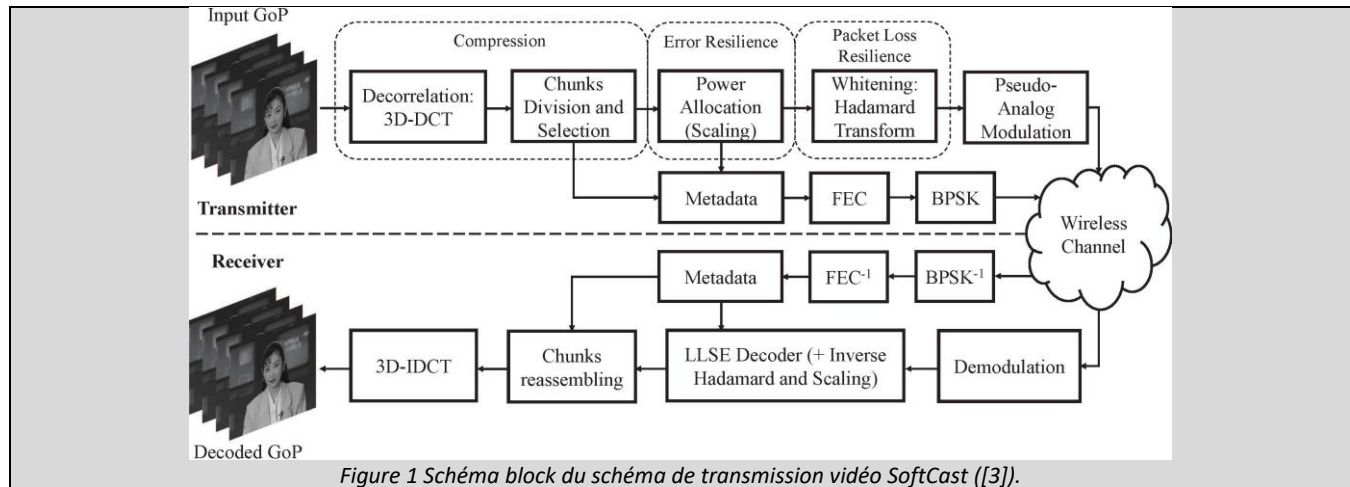
Récemment, un nouveau schéma *SoftCast* a démontré un potentiel élevé pour la diffusion de vidéo sur des réseaux sans fil avec des caractéristiques de canaux variant dans le temps, en comparaison aux schémas vidéo classiques (H.264/AVC, H.264/SVC, HEVC). Particulièrement utile en situation de Broadcast, il permet de traiter les problèmes rencontrés dans les systèmes de diffusion vidéo sans-fil conventionnels comme le *cliff-effect* [1]. L'objectif de ces travaux est de proposer une amélioration du schéma *SoftCast* originel, en effectuant un encodage adaptatif prenant en considération les contraintes liées à la bande passante ainsi que les caractéristiques des contenus vidéos. Nos résultats de simulations mettent clairement en évidence l'importance de s'adapter au contenu vidéo dans *SoftCast*. Selon les caractéristiques spatio-temporelles du contenu vidéo, une amélioration du PSNR de l'ordre de 2dB est observée pour les contenus vidéos à faible index spatial et temporel (SI, TI) [2], alors qu'un gain en terme de complexité d'encodage allant jusqu'à 40% est observé à qualité quasi-identique pour les contenus vidéos à haut SI, TI.

Mots-clefs (3 à 5 max.) :

Transmission vidéo sans-fil, SoftCast, Transmissions Vidéo non Codées, Index Spatio-Temporels

Contexte et état de l'art :

Différent des standards de transmission vidéo actuels, *SoftCast* est un système de codage conjoint source-canal où les pixels sont traités par des opérations linéaires successives et directement transmis sans quantification ni codage. De fait, il permet d'obtenir une dégradation progressive de l'image (graceful degradation [3]) : la qualité reçue est une fonction linéaire du bruit du canal [4]. Le schéma de base de *SoftCast* [3] est présenté dans la Fig. 1. *SoftCast* travaille sur des groupes d'images (GoP) et utilise une transformation DCT-3D. Les plans DCT résultants sont ensuite divisés en blocs rectangulaires de coefficients transformés appelés *chunks*. Ensuite, une mise à l'échelle des coefficients transformés (*scaling*), garantit que les données vidéo sont bien protégées contre le bruit des canaux. La transformée d'Hadamard est par la suite appliquée aux coefficients mis à l'échelle afin de fournir une résilience à la perte de paquets. Ce processus transforme les *chunks* en *slices*. Chaque slice est une combinaison linéaire de tous les *chunks*. Enfin, ces paquets sont transmis d'une manière pseudo-analogique à l'aide d'une modulation dense : Raw-OFDM [16]. Cette étape se fait en ignorant les codes correcteurs d'erreurs (FEC) et la partie de modulation classique d'un système OFDM. Ainsi, dans *SoftCast*, au lieu de débit binaire, le débit symbole est considéré. Etant donné que les coefficients transformés sont envoyés en paires (plan I et Q en OFDM), la bande passante maximale canal dans *SoftCast* peut être définie de la manière suivante : $BW_{max} = (résolution\ vidéo * frame\ rate) / 2$. Lorsque la bande passante de l'application ne permet pas la transmission de tous les coefficients transformés, *SoftCast* opère en jetant des *chunks*. Les *chunks* sont préalablement triés par ordre décroissant de variances énergétiques. Dans *SoftCast*, le niveau de compression CR est défini comme suit : $CR = BW_{ava} / BW_{max}$ où BW_{ava} dénote la bande passante disponible à l'émetteur. En parallèle du processus de transformation, l'émetteur *SoftCast* envoie des métadonnées. Ces données sont constituées de la moyenne et de la variance des *chunks* transmis, ainsi que d'une carte binaire, qui indique les positions des *chunks* ignorés dans le GoP. Ces métadonnées sont fortement protégées et transmises de manière robuste pour assurer une réception correcte. Du côté du récepteur, un décodeur LLSE (Linear Least Square Error) est utilisé afin d'obtenir la meilleure estimation des coefficients DCT reçus. En utilisant les métadonnées et les opérations inverses du décodeur, ces coefficients sont ensuite réassemblés pour reformer les plans DCT. Ces derniers sont ensuite transformés via un processus 3D-DCT inverse. Dans la configuration actuelle, le schéma *SoftCast* utilise une taille de GoP fixe qui ne permet pas de s'adapter aux spécificités spatiotemporelles de chaque contenu vidéo.



Travail proposé (Décrire l'objectif du travail et indiquer clairement le problème/défi technique étudié) :

Dans ces travaux, nous examinons tout d'abord l'impact du contenu vidéo sur la qualité reçue à la réception pour différentes tailles de GoP. Comme mentionné précédemment, le but est de proposer une amélioration du schéma *SoftCast* originel, en effectuant un encodage adaptatif prenant en considération les contraintes liées à la bande passante du canal ainsi que les caractéristiques des contenus vidéos. Ceci dans le but de proposer le meilleur compromis bande-passante canal/reconstruction vidéo. Le défi technique consiste à proposer un algorithme adaptatif pouvant s'inscrire dans un schéma de type *SoftCast* qui détermine à la volée la taille de GoP optimale suivant les variations d'activités spatio-temporelles pouvant avoir lieu dans les séquences vidéos.

Résultats :

La métrique du Peak Signal-to-Noise Ratio (PSNR) est tout d'abord utilisée comme indicateur objectif de la qualité vidéo reçue ceci pour différents rapports signal/bruit (CSNR) et différentes tailles de GoP. Le schéma *SoftCast* est implémenté comme dans [3]. Cependant, afin d'améliorer la qualité en réception, la moyenne de chaque image est soustraite et envoyée en metadata comme décrit dans [5]. Trois tailles de GoP sont ici utilisées : 8, 16 et 32 images. Compte tenu des limitations de place du résumé étendu, 2 séquences CIF parmi l'ensemble des tests effectués sont ici présentées : *Akiyo* et *Husky* (index spatio-temporels très hétérogènes). Pour une comparaison juste, un même bruit est généré et utilisé pour toutes les configurations.

Séquence		CSNR(dB)	0	5	10	15	20	25
Akiyo (SI = 9.57, TI = 5.86)	GoP 8	33,68	38,09	42,69	47,31	51,67	56,69	
	GoP 16	35,14	39,50	44,05	48,62	52,86	58,88	
	GoP 32	36,25	40,56	45,06	49,53	53,79	60,94	
Husky (SI = 20.96, TI = 62.02)	GoP 8	20,57	24,25	28,64	33,39	38,27	43,17	
	GoP 16	20,61	24,29	28,67	33,42	38,31	43,20	
	GoP 32	20,62	24,28	28,67	33,42	38,31	43,20	

Tableau 1. Tableau des PSNR pour différentes tailles de GoP et différents CSNR, CR = 1 (pas de compression).

Puisque le PSNR est une métrique basée sur l'erreur quadratique moyenne (MSE) n'étant pas corrélée avec la perception humaine, nous utilisons également l'index SSIM (Structural SIMilarity) en tant que seconde métrique :

Séquence		CSNR(dB)	0	5	10	15	20	25
Akiyo (SI = 9.57, TI = 5.86)	GoP 8	0,875	0,942	0,976	0,991	0,996	0,998	
	GoP 16	0,901	0,954	0,981	0,993	0,997	0,999	
	GoP 32	0,916	0,962	0,984	0,994	0,997	0,999	
Husky (SI = 20.96, TI = 62.02)	GoP 8	0,671	0,817	0,912	0,964	0,987	0,995	
	GoP 16	0,671	0,817	0,912	0,964	0,987	0,995	
	GoP 32	0,670	0,816	0,911	0,963	0,987	0,995	

Tableau 2. Tableau des index SSIM pour différentes tailles de GoP et différents CSNR, CR = 1 (pas de compression).

Les résultats présentés dans le tableau 1 et 2 montrent que :

- Pour les séquences vidéos présentant une faible activité spatiotemporelle (de type *Akiyo*), une GoP de grande taille permet d'obtenir une meilleure qualité de reconstruction.

- Dans le cas contraire (forte activité spatiotemporelle) comme *Husky*, il n'y a pas d'intérêt à augmenter la taille de la GoP en terme de qualité de reconstruction. Ceci est d'autant plus vrai que la complexité augmente selon $O(K \log(K))$ avec K le nombre de pixels dans un GoP [3]. Maintenir une taille de GoP faible permet dans ce cas une réduction de la complexité allant jusqu'à 40%.

La figure 2 illustre également les images reconstruites pour les différentes configurations dans le canal le plus perturbé correspondant à un CSNR égal à 0dB :

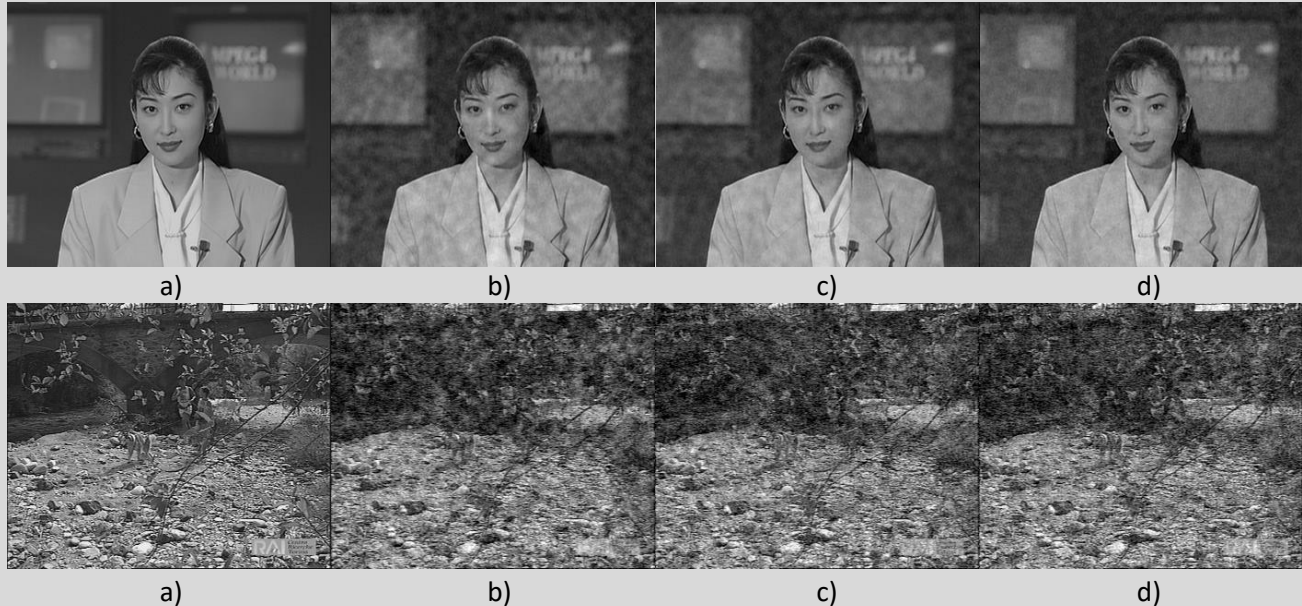


Figure 2 Comparaison visuelle de la qualité à la réception dans un canal à 0dB, CR = 1 (pas de compression). De haut en bas : Séquence Akiyo, Séquence Husky. De gauche à droite : a) première image de la séquence originale, b) première image reconstruite GoP = 8, c) première image reconstruite GoP = 16, d) première image reconstruite GoP = 32.

Des premiers tests concernant les applications à bande passante restreinte ont également été effectués. Le niveau de compression appliqué a été réduit, passant de 1 à 0.25 (25% des coefficients transmis). Les conclusions obtenues dans la première étude (CR = 1, pas de compression) sont également valables lorsque l'application visée requiert l'abandon de certains coefficients. Etant donné les contraintes de place du résumé étendu, ces résultats ne peuvent pas être intégrés ici et seront donc présentés lors de la présentation orale.

Conclusion et perspectives :

Les résultats préliminaires de simulations mettent clairement en évidence l'importance de l'adaptation au contenu vidéo dans un système de transmission vidéo sans fil *SoftCast*.

L'impact du contenu vidéo dans des environnements à bande passante restreinte est en cours d'étude afin de généraliser les travaux actuels. Une proposition de schéma de transmission vidéo de type *SoftCast* effectuant un encodage adaptatif est en cours. Cette version adaptative permet de prendre en considération à la fois les contraintes liées à la bande passante disponible mais aussi les caractéristiques des contenus vidéos et propose le meilleur compromis bande-passante/reconstruction vidéo. Selon les contraintes matérielles de l'application visée, un compromis qualité-complexité pourra être choisi permettant d'obtenir la meilleure qualité disponible pour la ressource matérielle utilisée.

Références (5 max.):

- [1] S. Kokalj-Filipovi, E. Soljanin, "Suppressing the cliff effect in video reproduction quality," Bell Labs Technical Journal, vol. 16, no. 4, pp.171–185, Mar. 2012.
- [2] ITU-T Recommendation, "Subjective video quality assessment methods for multimedia applications," 1999.
- [3] S. K. Jakubczak, "SoftCast: exposing a waveform interface to the wireless channel for scalable video broadcast," MIT PhD Thesis, 2011.
- [4] R. Xiong, F. Wu, J. Xu, X. Fan, C. Luo, W. Gao, "Analysis of decorrelation transform gain for uncoded wireless image and video communication," IEEE Trans. Image Process., vol. 25, no. 4, pp. 1820–1833, Apr. 2016.
- [5] A. Trioux, F.-X. Coudoux, P. Corlay, M. Gharbi, "A Comparative Preprocessing Study for SoftCast Video Transmission", Accepted in IEEE ISIVC Symposium, 27-30 Nov. 2018, Rabat, Morocco.



Titre :

Étude du mouvement d'un motif à l'aide d'un réseau de neurones impulsionnels

Auteurs :

Veïs OUDJAIL, Jean MARTINET

Résumé (100-200 mots) :

Depuis ces dernières années, les approches deep learning sont devenues incontournables dans le domaine de la vision artificielle, en s'illustrant par des résultats supérieurs aux méthodes précédentes. Malgré tout, elles possèdent quelques inconvénients : elles nécessitent, notamment, beaucoup de données ainsi que des ressources de calculs importantes. Pourtant, il existe de nombreuses tâches en vision où ces limitations deviennent des contraintes fortes, il suffit pour cela de s'intéresser aux applications s'exécutant dans un environnement embarqué, on peut citer, par exemple, le pilotage de drone.

Les réseaux de neurones impulsionnels peuvent être une réponse intéressante. En effet, leur apprentissage est non-supervisé, évitant ainsi l'étiquetage de nombreuses données. De plus, ces modèles se prêtent bien à des implémentations matérielles basse consommation dédiées.

Dans les travaux présentés ici, nous allons nous intéresser à une tâche élémentaire en vision, du fait de la complexité du modèle, à savoir pouvoir caractériser le mouvement d'un motif simple afin de comprendre les mécanismes permettant de les utiliser pour des opérations plus complexes.

Mots-clefs (3 à 5 max.) :

Analyse vidéo, Réseaux de neurones impulsionnels, Apprentissage automatique, Approche bio-inspirée.

Contexte et état de l'art :

Les réseaux de neurones impulsionnels (*Spiking Neural Networks*, SNN) [1] représentent une classe spéciale au sein des réseaux de neurones artificiels. Ils sont proches du modèle biologique, où les neurones communiquent par des séquences d'impulsions sortantes (spikes). La règle d'apprentissage, Spike-Timing Dependant Plasticity (STDP), est non supervisée.

Du fait de la nature du modèle, il faut être capable d'encoder ces données et de les transformer en une série temporelle d'impulsions. Une solution est de choisir le format *Address-Event Representation* (AER) pour représenter une séquence vidéo. Il encode de manière binaire, indépendamment pour chaque pixel, une variation positive (événements de type ON) ou négative (événements de type OFF) de luminosité. Certains capteurs bio-inspirés, comme les *Dynamic Vision Sensors* (DVS) représentent l'information de cette manière. Le caractère asynchrone des SNN se prête bien au traitement de l'information issue de ces capteurs : leur sortie peut être traitée et convertie pour nourrir la couche d'entrée du réseau.

L'usage des SNN en vision connaît un essor récent. Concernant la classification d'images, des datasets au format AER ont été développés pour faciliter leur évaluation, comme Poker-DVS ou MNIST-DVS [3]. En s'intéressant à l'analyse vidéo, il existe plusieurs tentatives d'application des SNN pour la classification de vidéo au format AER. Amir *et al.* [4] ont conçu un outil qui permet de reconnaître en temps réel 10 gestes de la main encodés dans ce format, avec une faible consommation énergétique, en exploitant un processeur neuromorphique de 1M neurones impulsionnels. Les résultats d'un tel système atteignent le taux de 96.5 % de réussite. Zhao *et al.* [5] combinent une architecture HMAX et SNN feed-forward afin de reconnaître les 3 actions humaines suivantes : se pencher, marcher et se lever/s'asseoir, dont les extraits de séquences font apparaître des mouvements diagonaux (se pencher), horizontaux (marcher) et verticaux (se lever/s'asseoir).

Bichler *et al.* [2] ont utilisé un SNN capable de reconnaître le mouvement d'une balle parmi 8 directions discrètes à partir des données de l'AER. Ils montrent également une application de comptage de voitures qui passent sur une voie d'autoroute.

Tous les travaux mentionnés précédemment utilisent les types d'événements ON et OFF. Avec la richesse qu'apportent ces informations, la direction de l'objet peut être déduite à un instant donné.

Prenons l'exemple d'une balle claire se déplaçant vers la droite sur un fond sombre. Le bord droit de la balle génère un ensemble d'événements ON et le bord opposé (gauche) génère simultanément un ensemble d'événements OFF. Dans cette configuration, le problème de reconnaissance de mouvement se réduit en un problème de reconnaissance statique de formes, car l'utilisation des deux types d'événements permet une reconnaissance *statique* : le positionnement spatial des événements renseigne sur la direction.

Par ailleurs, dans le cas où l'objet est plus grand que le champ de vision du capteur, les événements ON et OFF sont séparés dans l'espace, ou même séparés dans le temps (c'est-à-dire qu'ils ne se produisent pas simultanément), rendant presque impossible une telle approche statique pour inférer un mouvement.

Notre objectif est d'étudier l'utilisation des SNN feed-forward capables d'apprendre et de reconnaître la dynamique du mouvement. Par conséquent, nous faisons délibérément abstraction des types d'événements, et nous considérons un unique type d'événement représentant les variations aussi bien positives que négatives.

Le but est d'obtenir des détecteurs sensibles à certaines directions de mouvement; indépendamment de la configuration spatiale des événements ON et OFF. Pour ce faire, nous visons à identifier la structure d'un réseau minimal dont la couche de sortie est capable d'apprendre la direction d'une texture binaire donnée en mouvement à l'entrée.

Travail proposé (Décrire l'objectif du travail et indiquer clairement le problème/défi technique étudié) :

Nos travaux visent l'identification d'une architecture SNN minimale capable de reconnaître le mouvement d'un motif en déplacement, en utilisant des données visuelles codées dans la représentation AER. Pour ce faire, nous avons généré des données synthétiques avec 8 motifs se déplaçant dans 4 directions (NORD, SUD, EST, OUEST) dans une fenêtre de résolution 5x5 torique.

Nous avons conçu plusieurs variantes d'un réseau feedforward totalement connecté à une seule couche avec un nombre variable de neurones pour la couche de sortie. Les réseaux, dont les poids synaptiques sont initialisés aléatoirement, sont entraînés à l'aide de la STDP, en présentant les différentes classes pendant quelques secondes. A l'issue de l'entraînement, nous ciblons une situation où certains neurones réagissent à une classe particulière.

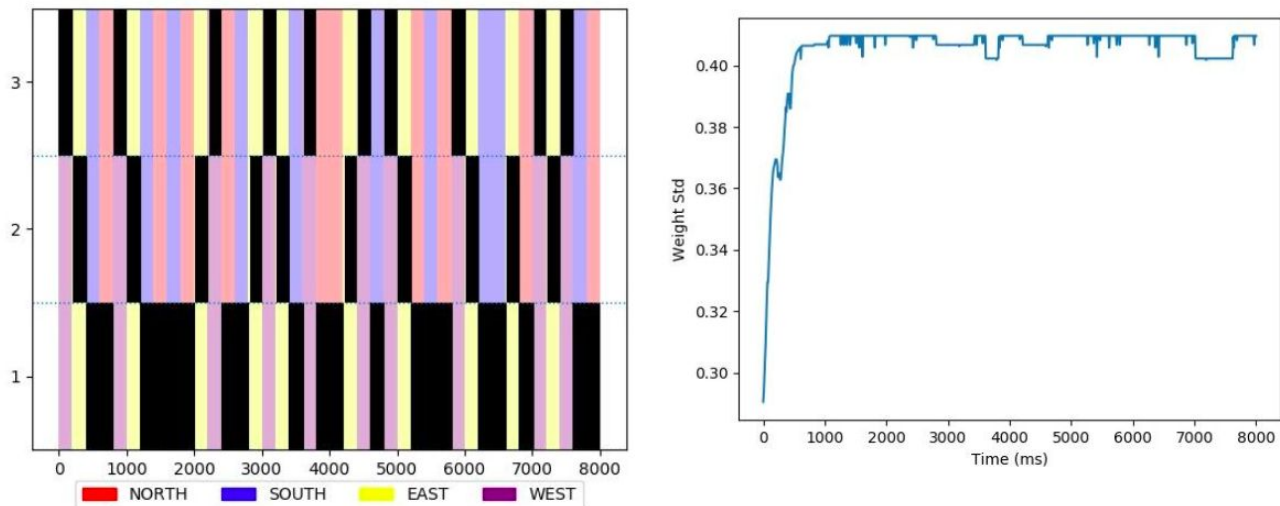
Résultats :

Les résultats montrent qu'après un court apprentissage, le réseau est généralement capable de reconnaître les directions présentées.

La figure de gauche transcrit l'activité d'un réseau doté de 3 neurones de sortie (ordonnées). L'axe des abscisses représente le temps de simulation en ms. Les barres verticales colorées indiquent la classe présentée (ordre aléatoire), et les marqueurs noirs indiquent les impulsions du neurone de sortie correspondant. Dans cet exemple, le motif présenté est une ligne verticale de 5 pixels de hauteur. Nous constatons que les classes d'entrée OUEST, EST et NORD/SUD sont reconnues par des neurones de sortie avec les indices 3, 2 et 1 respectivement - les directions NORD/SUD étant indiscernables du fait que la ligne occupe toute la hauteur de la fenêtre et que la fenêtre est torique. Il s'agit là d'un cas de reconnaissance parfaite, le succès étant variable selon les motifs et les paramètres choisis pour le modèle.

Afin de suivre l'évolution du système, nous avons tracé la valeur de l'écart-type des poids synaptiques (figure de droite), avec l'hypothèse que ces poids devraient, au cours de la simulation, s'écarter de leurs valeurs

initiales en se spécialisant : soit en tendant vers 0 - ce qui indique que le neurone est devenu insensible au stimulus, soit vers 1 - ce qui indique que le neurone est devenu très sensible au stimulus. La figure montre une forte augmentation au début de la simulation, indiquant la spécialisation des neurones, puis un plateau indiquant un état stable du réseau.



Conclusion et perspectives :

Nos principales constatations sont les suivantes : nous avons entraîné avec succès des SNN de petites tailles pour la reconnaissance de la direction du mouvement de motifs simples. Quelques millisecondes de stimulation non supervisée suffisent pour atteindre une situation stable, et la poursuite de la stimulation ne dégrade pas l'apprentissage.

Dans la suite de ce travail, nous allons nous intéresser à explorer l'espace des paramètres du modèle afin d'identifier les situations favorables à l'apprentissage et à la stabilité, ainsi qu'à la généralisation à des motifs quelconques. L'objectif de long terme est de pouvoir utiliser ces réseaux dans des applications réelles plus complexes comme la caractérisation du flux optique sur des données en provenance du capteur DVS ou d'autres capteurs bio-inspirés.

Références (5 max.):

- [1] H. Paugam-Moisy, S. Bohte. Computing with spiking neuron networks. Handbook of natural computing, 335-376, 2012.
- [2] O. Bichler, D. Querlioz, S.J. Thorpe, J.-P. Bourgoid, C. Gamrat. Extraction of temporally correlated features from dynamic vision sensors with spike-timing-dependent plasticity. Neural Networks (32):339-348, 2012.
- [3] T. Serrano-Gotarredona, B. Linares-Barranco. Poker-DVS and MNIST-DVS. Their History, How They Were Made, and Other Details. Front. Neurosci. (9:481), 2015.
- [4] A. Amir, B. Taba, D.J. Berg, T. Melano, J.L. McKinstry, C.D. Nolfo et al. A low power, fully event-based gesture recognition system. In CVPR, pages 7388–7397, 2017.
- [5] B. Zhao, R. Ding, S. Chen, B. Linares-Barranco, H. Tang (2015). Feedforward Categorization on AER Motion Events Using Cortex-Like Features in a Spiking Neural Network. IEEE Trans. Neural Netw. Learning Syst., 26(9), 1963-1978.



Titre :

Video summarization based on human action detection and recognition

Auteurs :

Omar Elharrouss, Noor Al-Maadeed, Somaya Al-Maadeed, Ahmed Bouridane, Azeddine Beghdadi

Résumé (100-200 mots) :

Video summarization based on human action detection and recognition is a challenging task for many applications including video surveillance, sports application, video indexing and Ambient Intelligence to name only few. The success of such operation highly depends on the quality of the detection, recognition and summarization of actions to be exploited in many other tasks. Therefore, it is deemed essential to develop efficient video summarization based on relevant features and descriptors of the visual content and especially those related to human action. In this work, an effective approach of action detection/recognition and summarization is proposed. Herein, the key idea is to detect and recognize actions of many people present in an observed scene. Our approach starts with a training phase based on the histogram-oriented gradient (HOG) of the binary image maps associated with each action of the training videos. Before starting the testing phase, the movement of persons in the observed scene are detected, tracked and a sequence of each human body silhouette is extracted. Each sequence contains many actions, thus; a function of separating each action as a shot is applied before computing the HOG. The HOG of each shot is then compared with HOGs of the training phase to recognize the action. Summarization of each action is performed for each detected person. The efficiency of the proposed approach is demonstrated through obtained results. Some perspectives and challenges are also discussed in this work.

Mots-clefs (3 à 5 max.) :

Action detection/recognition, HOG, Machine learning, Video summarization;

Contexte et état de l'art :

The increasing use of video surveillance systems and the huge volume of recorded videos make the extraction of the relevant information in real-time from this visual information very challenging problem. Although the techniques of information storage and retrieval are being rapidly developed, their efficient exploitation in video surveillance platforms is still hampered by the lack of smart solutions to better exploit these techniques in the case of large volume data. The extracted information is useful to identify and detect many events and assist to predict many events that usually happen in public scenes. Recently, a number of researchers focused their studies on finding effective techniques to summarize the useful information from video. Based on the semantics and the content of videos, an effective summarization can focus on one target like motion, abnormal events, and human action [1, 2]. In order to summarize human action some authors in [3] use the compact descriptors for visual search (CDVS). Local features and global structure of the image provided by CDVS features is a good technique for features extraction in real time because of the use of computing optimization. Dasari et al. classified actions by tracking CDVS features trajectories of the detected human body [4]. Qazi et al. propose an action recognition method using Scale Invariant Feature Transform (SIFT) and (HOG) features [5].

Travail proposé (Décrire l'objectif du travail et indiquer clairement le problème/défi technique étudié) :

Human action recognition can provide an effective information for many computer vision applications. It is also an important task for video analysis. Most of the proposed methods recognize the action of one person in the scene. Several methodologies have been proposed for the summarization of actions. In this work, a learning based method using multi-action recognition and summarization is proposed. In the training phase, we compute of Temporal Difference Map (TDM) of the binary images of each action from training datasets, after that the histogram of oriented gradient (HOG) of each result is calculated. The testing phase starts with the detection of human bodies within the scene, followed by tracking them to provide a sequence of each human body. Within each sequence, shots that represent the homogenous part of the sequence are extracted using the histogram of the similarities between the images of the sequence. After that, the same features in the training phase of each shot (sub sequence) are used to recognize which action is contained in it by comparing with all histograms computed in the training phase. After the recognition stage, the summarization of the action in the scene is conducted by

choosing one frame from each shot that represents an action. Figure 1 illustrates the flowchart of the proposed method.

Résultats :

In order to evaluate the performance of the proposed method, four action recognition datasets are utilized; including KTH, Weizmann and UCF-ARG datasets. In addition, PET09 dataset, that contains many persons in the scene, is used to test the proposed approach. Summarization of actions was tested on our generated video due to the unavailability of videos containing many actions. The generated HOGs in the testing phase are compared with those of training phase by computing the distance between the histograms. Obtained results demonstrate the accuracy of the proposed methods through combining temporal difference map and histogram of oriented gradient. The proposed method was compared with the state-of-art methods on the same datasets, and had a successful recognition rate of 97% (Table 1).

Conclusion et perspectives :

The summarization of human actions is an important task for many applications; for instance in sports applications it helps recognizing and understanding player poses thereby allowing judges to make good decisions in the case of player fouls. The existing approaches of action recognition are mainly restricted to a single person in a video sequence. In order to perform the recognition and summarization of multiple persons within a video sequence, the proposed approach extracts actions of each person present in the scene by detecting and tracking human bodies. Shots for each sequence are then detected in order to separate the action (action shot) within sequence. After that, each shot was used to recognize the action occurring using training features comparison (HOGs). After the recognition stage, the summarization of the action in the scene was made by choosing one frame from each shot that represents an action. The obtained results are convincing, and revealed that the algorithm can be used in real-time thanks to its simplicity and the reduced number of features.

Références (5 max.):

- [1] Zhang, S., et al. (2016). "Context-aware surveillance video summarization." IEEE Transactions on Image Processing 25(11): 5469-5478.
- [2] Lin, W., et al. (2015). "Summarizing surveillance videos with local-patch-learning-based abnormality detection, blob sequence optimization, and type-based synopsis." Neurocomputing 155: 84-98.
- [3] Kamiński, Ł., et al. (2017). Human activity recognition using standard descriptors of MPEG CDVS. Multimedia & Expo Workshops (ICMEW), 2017 IEEE International Conference on, IEEE.
- [4] Dasari, R. and C. W. Chen (2018). MPEG CDVS Feature Trajectories for Action Recognition in Videos. 2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), IEEE.
- [5] Qazi, H. A., et al. (2017). Human action recognition using SIFT and HOG method. Information and Communication Technologies (ICICT), 2017 International Conference on, IEEE..

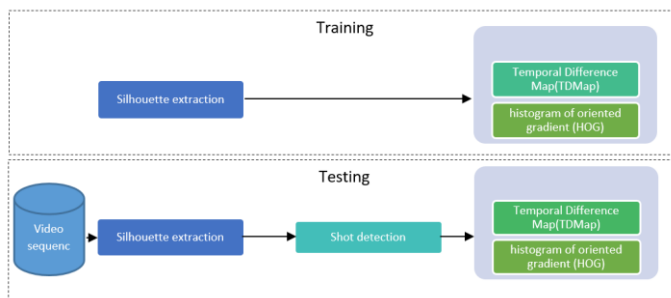


Figure 1: Flowchart of the proposed method

Table 1 Recognition rate comparison of KTH Weizmann and UCF-ARG Datasets

KTH dataset	
Method	Accuracy (%)
(Kamiński, Maćkowiak et al. 2017) [3]	92%
(Dasari and Chen 2018) [4]	87%
Proposed Method	98%
Weizmann dataset	
(Kamiński, Maćkowiak et al. 2017) [3]	81%
Proposed method	98%
UCF-ARG dataset	
(Qazi, Jahangir et al. 2017) [5]	90%
Proposed method	97%



Titre :

Estimation de la qualité biométrique d'échantillon de voix dans le cadre de la reconnaissance du locuteur

Auteurs :

CABANA Antoine

CHARRIER Christophe

Résumé (100-200 mots) :

L'avancée des techniques de reconnaissance du locuteur permettent d'obtenir des systèmes d'authentification n'admettant que des taux d'erreurs biométriques faibles. Néanmoins, ces performances sont généralement atteintes sur des bases de données biométriques capturées dans des conditions favorables. Cependant, l'environnement sonore influe de manière importante sur les performances observées sur ce type de systèmes biométriques. Ce travail se propose d'étudier la qualité d'échantillons vocaux en vue de leur utilisation en tant qu'échantillon de comparaison dans le cadre d'un processus de reconnaissance biométrique.

La méthodologie proposée se repose sur l'extraction de multiples caractéristiques afin d'estimer la qualité du signal du point de vue biométrique. Un entraînement de la métrique de qualité s'est fait par le biais d'une corrélation de scores issues de comparaison biométrique entre une référence originale, et des échantillons ayant subits différents niveaux de bruits additifs ; et un ensemble de caractéristiques extraites du signal audio. Les résultats obtenus montrent une corrélation forte entre la métrique entraînée et la qualité biométrique constatée.

Mots-clefs (3 à 5 max.) :

Reconnaissance du locuteur

Métrique de qualité

Traitement du signal

Qualité biométrique

Contexte et état de l'art :

L'estimation de la qualité des échantillons vocaux et/ou biométrique a déjà étudiée dans différents travaux de recherches. Ainsi Grother et Tabassi[1] proposent différentes recommandations et exigences concernant la définition d'une métrique de qualité biométrique, en particulier sur le facteur de forme de cette métrique. Ainsi, ce dernier se doit de se présenter sous la forme d'une valeur scalaire qui doit être corrélée positivement avec la qualité réelle d'un échantillon biométrique, et idéalement comprise entre 0 et 100. La nature de cette corrélation se doit d'être au moins de rang, une corrélation de proportionnalité étant idéale.

La qualité de la voix a principalement été étudiée afin d'estimer la qualité d'écoute d'un échantillon, ou son utilité dans le cadre d'un processus de reconnaissance biométrique.

L'estimation de la qualité auditive de la voix a fait l'objet d'un travail de normalisation de la part de l'Union Internationale des Télécommunications[2]. La métrique de qualité proposée est une méthode d'estimation objective et monodimensionnelle de la qualité de la voix. Pour ce faire, différentes familles de caractéristiques sont extraites et sont ensuite fusionnées afin d'estimer la qualité d'écoute de l'échantillon considéré.

Dans le cadre de la reconnaissance du locuteur, l'estimation de la qualité des échantillons biométriques est utilisée dans le processus de reconnaissance afin d'estimer la fiabilité d'une décision. Les travaux sur l'estimation de la qualité biométrique de la voix se reposent généralement sur la création d'un UBM (*Universal Background Model*, dans le cadre de la voix basé sur des mixtures de gaussiennes). Néanmoins, cette méthode impose une phase d'entraînement importante, puisque l'objectif d'un UBM est de capturer la nature du signal vocal, la base d'entraînement doit donc comporter de nombreux individus et avoir été capturée à l'aide de différents microphones.

Travail proposé (Décrire l'objectif du travail et indiquer clairement le problème/défi technique étudié) :

La méthode proposée vise à estimer la qualité d'un échantillon de voix afin d'estimer la qualité réelle de cet échantillon biométrique. Dans le cadre de ce travail, la qualité biométrique réelle ne pouvant être estimée de manière exacte, celle-ci a été approximée par les performances obtenues au cours d'une comparaison biométrique. Ainsi, les scores de comparaison légitimes sont utilisés en tant que vérité terrain afin de représenter son utilité au regard d'un processus de reconnaissance biométrique.

La méthode proposée se repose sur l'extraction de différentes familles de caractéristiques et sur leur fusion. Certaines caractéristiques

Les différentes caractéristiques extraites sont :

- La gigue et le scintillement[4], la gigue mesure et qualifie les variations successives de la fréquence fondamentale au cours du temps ; le scintillement mesure les variations de l'amplitude au cours de périodes successives.
- Les valeurs d'asymétrie et de kurtosis observés sur les coefficients MFCC et LPC : l'asymétrie et le kurtosis sont des mesures permettant de décrire les distributions observées pour les coefficients MFCC et LPC. Les coefficients MFCC (*Mel-Frequency Cepstral Coefficients*) sont utilisés de manière récurrente en tant que caractéristiques pour le traitement de la voix et particulièrement dans le cadre de la reconnaissance du locuteur. Les coefficients LPC (*Linear Predictive Coding* ou codage prédictif linéaire) représentent quant à eux des informations extraites de l'enveloppe spectrale d'un signal de voix.
- La probabilité d'activité vocale (VAD) : les algorithmes de VAD sont utilisés afin de déterminer la probabilité qu'une fenêtre temporelle contienne du discours. La mise en œuvre de détermination de la VAD est celle proposée par Sohn *et. al.*[5].
- L'estimée de la fréquence fondamentale est réalisée à l'aide de l'algorithme DYPSA (*cf.* Naylor *et. al.* [6]), qui permet de déterminer l'instant de fermeture glottale.
- Des mesures harmoniques : la rugosité et l'inharmonicité sont extraites du signal de discours, afin de respectivement caractériser la dissonance sensorielle et le ratio d'énergie en dehors du spectre harmonique idéal.
- La valeur efficace du signal permet d'estimer la puissance du signal sur l'ensemble des fenêtres temporelles extraites d'un signal vocal, l'ensemble des valeurs permet de quantifier la variation de puissance au fil des *frames*.
- L'indice de modulation du signal audio reflète la variation de l'amplitude au cours du temps. Ce dernier est particulièrement sensible à l'introduction de bruit additif dans un échantillon vocal.

Conformément aux recommandations formulées par Grother et Tabassi [1], une métrique de qualité doit se présenter sous la forme d'une valeur scalaire. Ainsi, l'ensemble des caractéristiques doivent ainsi être combinées afin d'obtenir une valeur scalaire, pour ce faire une combinaison linéaire est utilisée :

$$QM = \sum_{i=0}^n \alpha_i x_i$$

où α_i représente les coefficients de la combinaison linéaire, et x_i représente la valeur associée à une caractéristique.

La valeur issue de cette combinaison linéaire doit être corrélée à la qualité biométrique observée afin de garantir sa pertinence. L'apprentissage de cette fonction de corrélation s'est faite par le biais d'un algorithme génétique afin de déterminer le vecteur de coefficients α_i permettant de maximiser la corrélation entre la valeur de qualité extraite et la qualité observée, approximée ici par les scores de comparaison légitime.

Résultats :

Le corpus d'échantillon vocaux est issu de seize personnes différentes, et pour chacune de celles-ci huit échantillons. La base d'échantillons est constituée des échantillons des onze premières personnes, et la base de test est constituée des cinq autres individus du corpus d'échantillon. Il n'y a donc pas de recoupement entre les individus des bases d'apprentissage et de test. Afin de constituer une base comprenant différents niveaux de dégradation, plusieurs bruits ambiants ont été ajoutés aux échantillons vocaux de la base d'apprentissage à différentes valeurs de SNR (ou rapport de signal sur bruit). Ainsi, les

différentes valeurs de SNR utilisées sont -3dB, 0dB, 3dB et 6dB

Afin d'augmenter le nombre d'essais, un procédé de bootstrap a été appliqué à l'apprentissage par algorithme génétique. À l'issue de cette phase d'apprentissage, pour les différentes métrique de qualité candidate, la meilleure est parvenue à obtenir un score de corrélation de 83 %, le vecteur issu de la moyenne des vecteurs obtenus par bootstrap obtenant un score de corrélation de 78 %.

Conclusion et perspectives :

La méthodologie proposée s'est appliquée à étudier et à définir une métrique de qualité biométrique permettant d'obtenir une estimation de la qualité biométrique de la voix. Cette méthode obtient des performances intéressantes. Le choix des caractéristiques utilisées est motivé par leur capacité à discriminer certaines caractéristiques spécifique à la voix, ou certaines distorsions, néanmoins d'autres caractéristiques doivent être étudiées afin d'en améliorer les performances. La sélection de nouvelles caractéristiques est susceptible de détecter d'autres types de distorsions que le bruit additif qui sont susceptible d'affecter les échantillons de voix. En effet, la technique développée ici, s'intéresse particulièrement à la perturbation la plus courante : le bruit ambiant.

L'extraction de caractéristique est donc une approche intéressante permettant d'obtenir des résultats pertinents, néanmoins la définition d'un UBM est une perspective intéressante. En effet, ces derniers ont pour objectif de créer une représentation d'une modalité indépendante des individus, et en conséquence nécessitent la constitution d'une importante base d'échantillon. Ainsi, face à cette nécessité d'une large base d'échantillon, l'implémentation de cette méthode a été initialement écartée, malgré sa capacité a déterminée si un signal est plus ou moins distant du modèle théorique que représente l'UBM.

Références (5 max.):

- [1] P. Grother and E. Tabassi, "Performance of biometric quality measures," IEEE transactions on pattern analysis and machine intelligence, vol. 29, no. 4, pp. 531–543, 2007.
- [2] L. Malfait, J. Berger, and M. Kastner, "P. 563 the itu-t standard for single-ended speech quality assessment," IEEE Transactions on Audio, Speech, and Language Processing, vol. 14, no. 6, pp. 1924–1934, 2006
- [3] J. Villalba, A. Ortega, A. Miguel, and E. Lleida, "Analysis of speech quality measures for the task of estimating the reliability of speaker verification decisions," Speech Communication, vol. 78, pp. 42–61, 2016
- [4] J. P. Teixeira, C. Oliveira, and C. Lopes, "Vocal acoustic analysis–jitter, shimmer and hnr parameters," Procedia Technology, vol. 9, pp. 1112–1122, 2013
- [5] J. Sohn, N. S. Kim, and W. Sung, "A statistical model-based voice activity detection," IEEE signal processing letters, vol. 6, no. 1, pp. 1–3, 1999
- [6] P. A. Naylor, A. Kounoudes, J. Gudnason, and M. Brookes, "Estimation of glottal closure instants in voiced speech using the dyspa algorithm," IEEE Transactions on Audio, Speech, and Language Processing, vol. 15, no. 1, pp. 34–43, 2007.



Titre :

Blind stereopair quality evaluation based on statistics of binocular contrast features

Auteurs :

Yu Fan, Mohamed-Chaker Larabi, Faouzi Alaya Cheikh, Christine Fernandez-Maloigne

Résumé (100-200 mots) :

Blind stereoscopic/3D image quality assessment (SIQA) is still challenging due to the poor understanding of the binocular perceptual mechanisms of the human visual system (HVS). In this paper, we develop a blind stereopair quality metric based on statistics aggregation of monocular and binocular local contrast features. In particular, for left and right views, we first extract statistical features of the image gradient magnitude (GM) and Laplacian of Gaussian (LoG), describing the image local structures from different perspectives. Then, the monocular statistical features are then fused to derive the binocular features based on a linear combination model using LoG-response-based and image local-entropy-based weightings, independently. These weightings can effectively model the strength of the view dominance with respect to the binocular rivalry (BR) behavior. Subsequently, we compute the GM of difference map between left and right views. Finally, the BR-inspired and disparity-related binocular features in addition to the subjective scores were jointly used to construct a learned regression model obtained by the support vector regression (SVR) algorithm. Experimental results on three widely used 3D IQA databases show that our model achieve competitive performance compared to state-of-the-art SIQA models.

Mots-clefs (3 à 5 max.) :

Blind Stereoscopic/3D image quality assessment, GM, LoG, binocular rivalry, support vector regression

Contexte et état de l'art :

With the recent advances in hardware equipment, the immersive stereoscopic/3D multimedia technologies have made great progress and thus been increasingly commercialized in entertainment and medical industries. This paper focuses on no-reference (NR) stereoscopic IQA (SIQA) to be close to real application where the reference images are unavailable. Previous studies on SIQA showed significant success using cyclopean image/feature, which is generated by fusing the left and right views with a binocular combination model and different weights. The cyclopean-based approaches require the disparity map, which is not always available, and its estimation is often error-prone and costly. To avoid these constraints, some blind SIQA algorithms predict the 3D quality without disparity/depth information, and thus based on BR-inspired combination of monocular 2D images features.

In this paper, we propose a new blind stereopair QA model based on statistics of both BF-inspired combined 2D monocular contrast and 3D disparity-related contrast which describe the image primitive structures. In particular, we first compute the image gradient magnitude (GM) and Laplacian of Gaussian (LoG) response for each single-view. Then, following the strategy described in [4], we extract the monocular joint statistics of the normalized GM and LoG features. Next, these monocular statistics are combined based on a linear summation model and different weights depending on strength of the views dominance, which model the BR behavior of the HVS. The weights are determined by LoG and local entropy (LE) maps of both views. Additionally, we compute the GM-based statistics of difference map between left and right views, which corresponds to disparity/depth (i.e., 3D) information. Finally, the combined 2D monocular and 3D binocular statistical features are jointly used to construct a learned regression model with support vector regression (SVR) algorithm. We comprehensively evaluate the performance of our proposed model, and achieve an extensive performance comparison with other competitive SIQA metrics on three widely used 3D image quality databases.

Travail proposé (Décrire l'objectif du travail et indiquer clairement le problème/défi technique étudié) :

Relying on the success of [1] on 2D QA, we use image GM and LoG responses to describe the image structural information from different perspectives. GM map shows the strength of local luminance variation, whereas LoG response measures the local luminance contrast after smoothing the noise.

Fig. 1 shows the framework of the proposed blind SIQA model. In fact, the Gaussian derivative functions can model the receptive field responses of neurons along the visual pathway. Therefore, we first compute the GM and LoG maps of left and right views independently to describe the primitive structures of 3D image. Here, we compute the GM and LoG maps using the first and second order derivatives of a circularly symmetric 2D Gaussian function G .

To develop a reliable stereopair quality metric, we account for 3D (depth) information in addition to monocular LoG and GM responses.

For a NR SIQA method, the ground truth disparity map is not available and its estimate is probably error-prone. Therefore, we propose to use the absolute difference map I_d between the left and right views to obtain the disparity. Next, we further compute GM_d of I_d in order to describe the binocular structures affecting 3D quality judgment of the HVS. Subsequently, a joint adaptive normalization procedure is applied to normalize both GM and LoG maps of both views so as to reduce the dependency of image statistics on the image content. For I_d , only the GM is computed. For each view, we then compute the joint statistics of the normalized GM and LoG maps by using both marginal and independent distributions to further extract efficient structural statistical features. Specifically, we derive the histogram-based feature vector (i.e., $\hat{F}_l$ or $\hat{F}_r$) for each view by concatenating their marginal and independent vectors. For disparity-related I_d , we only use the marginal distribution (i.e., $\hat{F}_d$) for GM responses to learning the quality prediction model.

We assume that 3D image quality is highly related to combination of monocular features distributions. Thus, we propose to combine the above-mentioned 2D monocular contrast statistics into binocular features considering the BR behavior of the HVS. We generate the binocular features $\hat{F}_b$ by a linear summation of the left and right features using weights estimating the strength of the view dominance. Here, we use both LoG and image local entropy (LE) maps to jointly describe the stimulus strength.

Given a test stereopair, we obtain 70 statistical features, among which 40 features for marginal and independent distributions of GM and LoG responses with LoG-based BR modeling, 20 features for marginal distributions of GM and LoG responses with respect to LE-based BR modeling, and 10 features for marginal distributions of GM responses with respect to disparity information. For the training stage, we adopt the SVR algorithm to map the abovementioned 70 features to the subjective 3D quality scores. In particular, we employ in this paper ϵ -SVR to generate the learned regression model with the kernel radial basis function. Finally, we use the learned regression model to predict 3D quality scores for the testing stage.

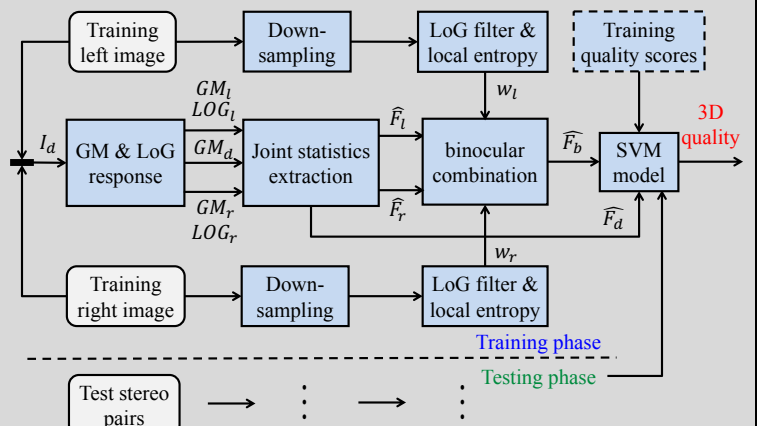


Fig. 1 : pipeline of the proposed blind SIQA model

Résultats :

Table 1. Overall performance of the SIQA models on LIVE 3D phase II, Waterloo IVC phase I and II databases. The ranking first and second model for each criterion are highlighted with red bold and blue bold texts, respectively.

SIQA model	LIVE 3D II			Waterloo IVC I			Waterloo IVC II		
	PCC	SROCC	RMSE	PCC	SROCC	RMSE	PCC	SROCC	RMSE
UQI	0.864	0.842	5.677	0.753	0.640	10.360	0.679	0.591	14.056
VIF	0.839	0.819	6.182	0.839	0.801	8.571	0.773	0.737	12.150
GMSD	0.843	0.827	6.073	0.748	0.741	10.439	0.742	0.758	12.836
BRISQUE	0.785	0.770	6.989	0.828	0.814	8.832	0.814	0.814	11.114
Chen-FR	0.910	0.905	4.693	0.657	0.510	11.910	0.613	0.578	15.740
Wang	0.916	0.916	-	0.930	0.918	-	0.892	0.869	-
Geng	0.921	0.919	5.400	0.864	0.810	9.469	-	-	-
Khan	0.932	0.922	-	0.934	0.925	-	0.910	0.905	-
Chen-NR	0.895	0.880	5.102	0.938	0.922	5.237	0.929	0.922	6.439
Fezza	0.925	0.908	3.018	0.904	0.898	-	0.890	0.886	-
Liu-2017	0.936	0.931	3.959	0.956	0.937	4.960	0.922	0.911	7.377
Liu-2018	0.932	0.930	4.120	0.948	0.924	5.567	0.941	0.936	6.373
Proposed	0.933	0.929	4.081	0.963	0.967	3.786	0.952	0.942	5.668

Table 2. Performance of the SIQA methods on symmetrically (i.e., S) and asymmetrically (i.e., As) distorted 3D images on LIVE 3D II, Waterloo IVC I and II databases. The ranking first and second method for each criterion are highlighted with red bold and blue bold texts, respectively.

SIQA model	LIVE 3D II				Waterloo IVC I				Waterloo IVC II			
	PCC		SROCC		PCC		SROCC		PCC		SROCC	
	S	As	S	As	S	As	S	As	S	As	S	As
UQI	0.941	0.795	0.939	0.755	0.814	0.724	0.635	0.631	0.759	0.629	0.565	0.566
VIF	0.928	0.777	0.916	0.732	0.918	0.788	0.913	0.755	0.861	0.719	0.830	0.687
GMSD	0.937	0.775	0.923	0.755	0.951	0.641	0.934	0.676	0.951	0.638	0.940	0.677
BRISQUE	0.857	0.700	0.849	0.663	0.925	0.786	0.891	0.787	0.929	0.748	0.900	0.772
Chen-FR	0.940	0.878	0.927	0.859	0.869	0.592	0.774	0.442	0.848	0.634	0.768	0.567
Wang	0.937	0.898	0.923	0.902	0.964	0.929	0.948	0.910	0.938	0.880	0.905	0.848
Khan	-	0.909	-	0.886	-	0.935	-	0.923	-	0.915	-	0.905
Chen-NR	-	-	0.918	0.834	-	-	0.934	0.907	-	-	0.944	0.899
Fezza	0.935	0.957	0.928	0.892	0.910	0.882	0.902	0.869	0.914	0.845	0.915	0.804
Liu-2017	-	-	0.933	0.905	-	-	0.967	0.934	-	-	0.917	0.904
Liu-2018	-	-	0.940	0.904	-	-	0.962	0.921	-	-	0.958	0.929
Proposed	0.945	0.926	0.925	0.909	0.967	0.968	0.980	0.960	0.961	0.934	0.958	0.922

Table 3. Performance of the SIQA methods for different types of distortion on LIVE 3D II database. The ranking first and second method for each criterion are highlighted with red bold and blue bold texts, respectively.

SIQA model	JPEG		JP2K		WN		Blur		FF	
	SROCC	RMSE	SROCC	RMSE	SROCC	RMSE	SROCC	RMSE	SROCC	RMSE
GMSD	0.887	3.223	0.868	4.831	0.934	3.873	0.921	3.811	0.935	3.811
VIF	0.778	4.146	0.822	5.462	0.822	5.907	0.951	2.223	0.934	3.700
UQI	0.739	4.412	0.831	5.314	0.840	5.622	0.954	2.763	0.956	3.181
Chen-FR	0.843	3.865	0.814	5.562	0.940	3.368	0.908	3.747	0.884	4.966
Shao	0.834	3.854	0.854	4.757	0.933	3.832	0.924	4.573	0.941	3.897
Voo	0.737	2.152	0.895	3.497	0.933	5.181	0.879	4.257	0.905	6.961
Wang	0.826	4.095	0.773	6.205	0.953	3.110	0.770	8.648	0.831	5.721
Chen-NR	0.867	3.342	0.867	4.298	0.950	3.531	0.900	4.725	0.933	4.180
Appina	0.839	4.756	0.864	5.087	0.932	4.325	0.846	6.662	0.860	6.519
Jiang	0.889	3.518	0.909	4.021	0.873	4.184	0.865	3.768	0.887	4.681
Liu-2018	0.858	3.068	0.908	4.022	0.940	3.536	0.901	2.570	0.924	3.879
Proposed	0.897	2.999	0.880	3.863	0.946	3.083	0.895	2.518	0.940	3.347

We evaluated the performance of the proposed blind SIQA model and the state-of-the-art models on three publicly available 3D image quality databases associating with differential mean opinion scores (DMOS): LIVE 3D IQA phase II database, and Waterloo IVC 3D phase I and II databases.

Table 1 shows the overall performance of the SIQA methods in terms of three criteria including PCC, SROCC and RMSE values on three databases. Our proposed model outperforms the other models on Waterloo IVC I and II, and delivers competitive performance and even better than FR-SIQA models on LIVE 3D II.

As asymmetrically distorted stereopairs are more challenging for quality prediction, we further evaluated the performance of the SIQA model on separated symmetric and asymmetric distortions of each database, and showed the results in Table 2. We observe that the proposed metric significantly outperforms most other SIQA metrics on Waterloo IVC I and II. For LIVE 3D II.

Besides, we compared the performance of the proposed models with other SIQA models for individual distortion types on LIVE 3D II on Table 3. The proposed model achieves high performance for JPEG and WN and FF distortions, and yields promising results for JP2K and Blur distortions.

Conclusion et perspectives:

In this paper, we present a blind SIQA model using the binocular histogram-based features from joint statistics of GM and LoG responses. The SVR algorithm with radial basis function kernel is then used to learn the prediction model, and finally to predict the perceived 3D quality. This work includes two main contributions: 1) the use of LoG responses and local entropy contrast on single/multi-scale left and right views to compute the weights of a linear summation model, and 2) accounting for structural features of disparity information. We provide an extensive evaluation showing high correlation with human judgment on widely used databases.

Références (5 max.):

- [1] W. Xue, X. Mou, L. Zhang, A. C. Bovik, and X. Feng, "Blind image quality assessment using joint statistics of gradient magnitude and Laplacian features," IEEE Trans. Image Process., vol. 23, no. 11, pp. 4850–4862, 2014.



Titre :

Une méthode originale basée sur l'IRM pour évaluer le phénotype infiltrant des gliomes de bas grade

Auteurs :

C. Brzenczek, Y. Gaudeau, S. Mezières, T. Obara, M. Blonski, F. Rech, L. Taillandier, JM. Moureaux.

Résumé :

Le gliome de bas grade est une maladie encore mal comprise et létale sur le long terme. Outre l'évolution de son volume, la forme de la tumeur joue également un rôle important sur le pronostic vital du patient. En effet, cette dernière peut limiter les possibilités thérapeutiques. Or, il n'existe aujourd'hui toujours pas d'outil utilisable dans le cadre clinique pour observer la forme du gliome et son évolution. Nous proposons de développer une méthode pour pallier ce manque. Ainsi, l'évaluation et la prédiction de l'évolution de la morphologie tumorale réalisée sur un grand jeu de données permettront d'améliorer la stratégie de soins.

Mots-clefs :

Gliomes de bas grade ; Imagerie par Résonance Magnétique ; Segmentation ; Reconstruction ; Infiltration

Contexte et état de l'art :

Les gliomes sont les tumeurs cérébrales primitives les plus fréquentes. Les gliomes diffus de bas grade (Gliomes de grade II de l'Organisation Mondiale de la Santé) infiltrent le tissu cérébral. Ils présentent une évolution lente, mais mènent inévitablement au décès du patient par un phénomène de transformation anaplasique. Le suivi de ces tumeurs s'effectue par imagerie par résonance magnétique (IRM, pondération T2 FLAIR), et les traitements possibles sont la chirurgie, la chimiothérapie et la radiothérapie.

Lors de leur croissance, on constate que ces gliomes peuvent prendre la forme d'une sphère refoulant partiellement le cerveau (dite « bulky »), ou une forme moins bien limitée, avec différents prolongements, témoignant d'un aspect plus infiltrant. Dans tous les cas, la tumeur est présente au-delà des limites visibles sur l'IRM, mais, cette infiltration est nettement moindre dans le cas des tumeurs « bulky ». Le type de morphologie prise par la tumeur conditionne alors l'efficacité des traitements dits focaux, comme la chirurgie ou la radiothérapie, puisqu'ils traiteront plus de tumeur et la récurrence sera donc observée plus tardivement.

Il est cependant difficile pour le praticien d'apprécier l'évolution tumorale d'un point de vue morphologique d'une IRM à l'autre (soit entre 3 et 6 mois d'intervalle). De plus, il n'existe actuellement pas d'outils en clinique qui permet d'apprécier visuellement la morphologie de telles tumeurs.

Des travaux récents ont étudié la sphéricité de la partie kystique de la tumeur dans le cas du glioblastome [1], ou encore la caractéristique « elliptique » d'un gliome de bas grade [2]. Cependant, il n'existe pas d'étude applicable au contexte de la prise en charge clinique du gliome diffus de bas grade.

Travail proposé :

Nous proposons ici une méthode originale qui permet de détecter, de quantifier, et de suivre l'évolution du phénotype infiltrant des gliomes de bas grade, permettant de discriminer les tumeurs « bulky » des tumeurs infiltrantes, et le passage d'un phénotype à un autre.

Pour ce faire, nous proposons de travailler sur les contourages réalisés par le médecin, lors des consultations de suivi, sur les IRM en séquence FLAIR du patient. Ces contourages sont utilisés pour reconstruire en 3D le volume tumoral sous MATLAB (2017).

On calcule dans un premier temps le barycentre de la masse tumorale comme étant la moyenne des coordonnées des voxels dans les 3 dimensions X, Y et Z (Voir figure 1). Après quoi, le volume de la tumeur est calculé par la somme du volume de chaque voxel. Enfin, le rayon de la sphère de volume équivalent est calculé selon la formule $R = \sqrt[3]{\frac{3V}{4\pi}}$, où V est le volume et R le rayon de la sphère.

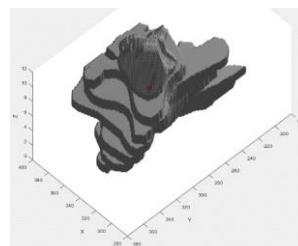


Figure 2 Reconstruction de la tumeur en 3D et de son barycentre (en rouge)

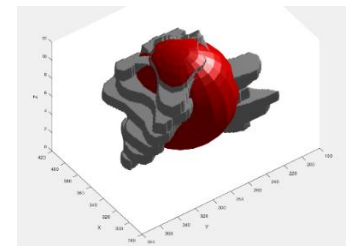


Figure 1 Reconstruction de la tumeur en 3D avec la sphère de volume équivalent centrée sur le barycentre

(Issu de la formule $V = \frac{4}{3} \pi * R^3$) La sphère est ensuite reconstruite, centrée sur le barycentre tumoral (Voir figure 2). Une partie de la tumeur se retrouve donc dans cette sphère fictive, et une autre partie à l'extérieur. Alors qu'une tumeur « bulky » concentrera une grande partie de son volume dans la sphère de volume équivalent, les tumeurs très infiltrantes montrant une morphologie allongée ou étoilée présenteront une plus grosse part de leur volume à l'extérieur de cette sphère (Voir figure 3).

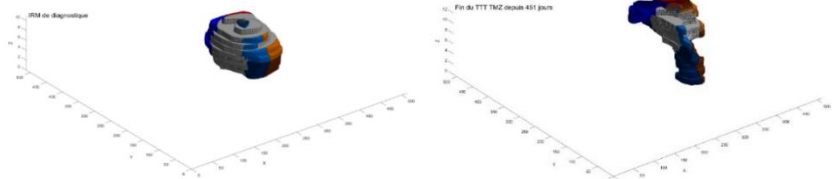


Figure 3 Exemple de reconstruction de deux tumeurs. Partie grise : Volume intra-sphère équivalente. Parties colorées : Volume extra-sphère équivalente découpé en 8 sous parties. A gauche : Tumeur « bulky », à droite : Tumeur infiltrante.

Résultats :

Le taux de volume intra-sphère équivalente (Volume tumoral intra-sphère / Volume tumoral total) décrit de manière quantitative l'éloignement de la tumeur d'une morphologie sphérique.

Plus encore, utilisant les contourages successifs, nous avons évalué la variation de ce taux de volume intra-sphère équivalent, et donc la capacité de la tumeur à évoluer vers un phénotype « bulky », ou au contraire, vers un phénotype plus infiltrant, chez trois patients.

Sur les trois patients, une tendance à l'infiltration au cours du temps est mise en évidence. En effet, la variation du taux de volume intra-sphère équivalent est négative. Ces résultats préliminaires mettent en évidence l'infiltration de la tumeur, ce qui est une première étape intéressante. (Voir Tableau 1)

Nous avons pu observer un volume tumoral stable pour le patient A, mais avec l'apparition d'un phénotype infiltrant pouvant témoigner d'une réponse dissociée au sein de la tumeur pendant la chimiothérapie, avec une partie « bulky » qui régresse et une autre, distante du barycentre, qui progresse. Ce phénomène non décrit à ce jour n'est pas évaluable en pratique clinique avec les méthodes actuelles, car consistant à évaluer le volume tumoral dans son ensemble, qui peut donc rester stable sans témoigner de la complexité et de l'hétérogénéité tumorale.

Patient	Variation max du taux en fonction de la durée	Variation Taux / Durée (pente)	Transformation anaplasique (TA) (Oui/Non) Durée Diagnostic à TA
A	-23 % en 1530 jours	-0,015	OUI, 3474 jours
B	- 6,5% en 1343 jours	-0,0048	OUI 1860 jours
C	- 4% en 2244 jours	-0,0018	NON (3893 jours Diag-Aujourd'hui)

Tableau 1 Variation du taux de volume intra-sphère équivalente, pente de la variation, présence de transformation anaplasique et délai de transformation par patient

Conclusion et perspectives :

Dans cet article, nous avons proposé une méthode originale qui permet d'évaluer le phénotype infiltrant d'une tumeur. Un tel outil n'existe aujourd'hui toujours pas dans le contexte de la prise en charge clinique. On ne connaît pas les facteurs génétiques susceptibles d'influencer l'évolution vers le caractère infiltrant des gliomes de bas grade. Cette méthode sera appliquée à la base de données de gliomes de bas grade du CHRU de Nancy. L'étude nous permettra de mieux cerner l'évolution du caractère infiltrant des gliomes de bas grade, afin de pouvoir émettre des hypothèses qui seront à tester en vue de la création d'un modèle de classification, et plus individuellement, de croissance, et d'inhibition de la croissance tumorale sous traitement.

Cette méthode pourrait permettre d'identifier des modifications de morphologie tumorale à ce jour non prise en compte dans la décision thérapeutique car non évaluables avec les outils actuels. Il serait alors possible de mettre en place de nouveaux algorithmes décisionnels quant aux traitements à proposer. L'ensemble de ces connaissances pourrait à terme considérablement améliorer la prise en charge thérapeutique du patient.

Actuellement cette méthode s'applique aux patients n'ayant pas pu obtenir une chirurgie. Il est envisageable d'appliquer la méthode aux patients opérés, une fois que la récurrence tumorale a dépassé un certain volume, dont le seuil reste à définir.

Références :

- [1] Mazurowski MA, Clark K, Czarnek NM. Radiogenomics of lower-grade glioma: algorithmically-assessed tumor shape is associated with tumor genomic subtypes and patient outcomes in a multi-institutional study with The Cancer Genome Atlas data. J Neurooncol. 2017; 133 (Issue1): 27-35
- [2] Pérez-Beteta J. Tumor Surface Regularity at MR Imaging Predicts Survival and Response to Surgery in Patients with Glioblastoma. Radiology. 2018; 288: 218-225



Titre :

Image Registration for Biomedical Images on Irregular Grids

Auteurs :

Nhung Pham, David Helbert, Pascal Bourdon, Philippe Carré

Résumé (100-200 mots) :

Microendoscope with spiral scanning produces images on irregular grids [5]. We propose a novel image registration strategy on irregular grids using Graph Wavelet (GW) constraint under the Log-Demons framework. The GW transformation is flexible to apply to different types of data regardless of the data point densities and how complex the data structure is. We also show how the Log-Demons framework can be adapted to optimize the objective function defined for images on irregular grids. Our results compare the performance of our method on irregular grid with the existing methods on a regular grid. To the best of our knowledge, this research is the first study on image registration on an irregular grid.

Mots-clefs (3 à 5 max.) :

Irregular grid, Graph Wavelets, microendoscopic images, Log-Demons

Contexte et état de l'art :

Our work focuses on registering two images acquired on irregular grids. Such an image has floating point coordinates and different point densities. The images are taken by a multiphoton microendoscopic whose optical fiber scans along a spiral path, starting from the center to the periphery of the area [5]. Such scanning produces a list of raw pixels on an irregular grid with floating coordinates.

Motivated by the fact that the system requires a robust image mosaics, it is important to have a sufficient non-rigid registration between the overlapping regions of two frames to compensate for the non-rigid deformations. These overlapping regions are at different positions of the two images, thus, have different point densities. Apart from the difficulties of the image grid characteristic, such images come up with much noise, illumination changes and non-rigid deformations.

The Log-Demons registration method [1] is known as a popular method for medical image registration, which is simple to implement and fast to compute. However, the optimization with a Gradient based update is easy to fall into local minima and lacks geometric invariance. It was shown in [3] that the graph spectrum is geometric invariant can be successfully embedded in a direct feature matching method under the Log-Demons framework to capture large and complex deformations. However, the registration using graph spectrum is sensitive to noise.

Recent research of Hammond et al [2] proposed a new computation of Spectral Graph Wavelets defined in the graph spectral domain. It is a flexible model constructed on an irregular weighted graph, thus, allows us to utilize the benefits of the graph representation of image on an irregular grid. More importantly, the Spectral Graph Wavelets not only represent images at different frequencies but also are isometric invariant under deformations of shape. Our previous work [4] proves the advantages of registration on regular grid using Spectral Graph Wavelets (SGW) [2] over the benchmark method that uses the graph spectra, especially when noise exists. In this paper, we exploit the previous research on image registration on regular grid that uses SGW under the Log-Demons framework to formalize a novel image registration method for images on irregular grids.

Travail proposé (Décrire l'objectif du travail et indiquer clairement le problème/défi technique étudié) :

Problem: Find transformation s from target image I_T and source image I_S such that $I_T = I_{S \circ s}$ with $I_T \in \mathbb{R}^M, I_S \in \mathbb{R}^N$ with floating point coordinates: $X_T = (V_{xt}, V_{yt}), X_S = (V_{xs}, V_{ys})$, resulting in different graph topologies (Fig. 1).

Background:

Graph definition: A graph $G = (\mathcal{V}, \mathcal{E})$ is composed of a finite set of vertices $\mathcal{V} = \{v_1, v_2, \dots, v_N\}$ and a set of edges $\mathcal{E}$ that represents the connection between edges, e.g. $(i, j) \in \mathcal{E}$ if v_i is connected to v_j . In this research, each sample point is represented by a vertex and edges represent neighboring structure of points. Such graph can be represented on an adjacency matrix A where A_{ij} undirected edge weight between vertices v_i and v_j .

The Laplacian matrix of a graph G is given by $L = D - A$, where D is a diagonal degree matrix and each value D_{ii} on the diagonal is the sum of the weights of all the edges incident to vertex v_i .

Graph Fourier Transform: Given a graph signal $f \in \mathbb{R}^N$ where N is the number of vertices, the eigendecomposition of the Laplacian matrix is written as $v_\ell = \lambda_\ell v_\ell$, where the set of eigenvalues and eigenvectors $\{\lambda_\ell, v_\ell\}_{\ell=1, \dots, N}$ form the eigensystem of the Laplacian matrix. The graph Fourier transform is given by:

$$\hat{f}(\ell) = \sum_{n=1}^N v_\ell^*(n) f(n),$$

Spectral Graph Wavelet Transform: Once the graph Fourier transform is defined, the Spectral Graph Wavelets are modulated by a generating kernel g . The Wavelet coefficients of graph signal f at scale t is given by:

$$W_f(t, n) = \sum_{\ell=1}^N g(t\lambda_\ell) \hat{f}(\ell) v_\ell(n)$$

In order to encode the low frequency of signal f , a second class of waveforms h is defined similarly to the low pass filter in classical wavelet transform, such that: $h(0) > 0$ and $\lim_{n \rightarrow \infty} h(x) = 0$. The scaling function coefficient $S_f(n)$ is given by:

$$S_f(n) = (h(L)f)(n) = \sum_{\ell=1}^N h(\lambda_\ell) \hat{f}(\ell) v_\ell(n)$$

Registration Method: I_S is first interpolated to the grid of I_T with scatter interpolation. Then, compute the SGWs of I_T and I_S on an irregular grid at t scales, resulting $t + 1$ column vectors for each image.

We choose the first k vectors at lower bands for registration, given by $W_T \in \mathbb{R}^{M \times k}$ for the target image and $W_S \in \mathbb{R}^{N \times k}$ for the source image. The registration, thus, can be cast into a feature matching problem that matches $\mathbf{T} = (\alpha_i I_T, \alpha_s X_T, \alpha_g W_T)$ and $\mathbf{S} = (\alpha_i I_T, \alpha_s X_T, \alpha_g W_T)$, where $\alpha_i, \alpha_s, \alpha_g$ account for the intensity, spatial and geometric weights. The objective function is defined as:

$$E(T, S, s) = \alpha_i \|I_T - I_{S \circ s}\|^2 + \alpha_s \|X_T - X_{S \circ s}\|^2 + \alpha_g \|W_T - W_{S \circ s}\|^2$$

The optimization is an iterative approach under the Log-Demons framework [1].

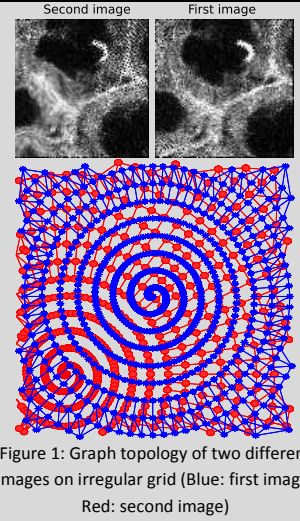


Figure 1: Graph topology of two different images on irregular grid (Blue: first image, Red: second image)

Résultats :

We benchmark our method against the two methods on regular grid (Spectral Demons, SGW Demons) and Spectral Demons on irregular grid.

We experiment two random pairs of frames in the sequence. The target and source images are cropped from the chosen frames. The system allows us to know the global shift between every two frames. For visualization, the images are projected on 100x100 Cartesian grid, and so for the tests on regular grid.

In Fig. 2a, each row represents the registered images of one pair of images with different methods. It can be clearly seen that the registration on a regular grid suffers from the smoothing effect.

Fig. 2b shows the Signal to Noise Ratio (SNR) values w.r.t the number of iterations. In both cases, the registration on regular grid is inconsistent. On irregular grids, the SGW Demons improves approximately 0.6 dB in the first pair and 0.5 dB in the second pair.

Fig. 3 allows us to see more clearly the result of our method applied on the first pair of images on irregular grids. In the representation of the difference, the target image is in the red channel, the source and registered images are in the green channel. The region covered by red line shows that the differences between the registered image and the target image have an improvement compared to the initial difference.

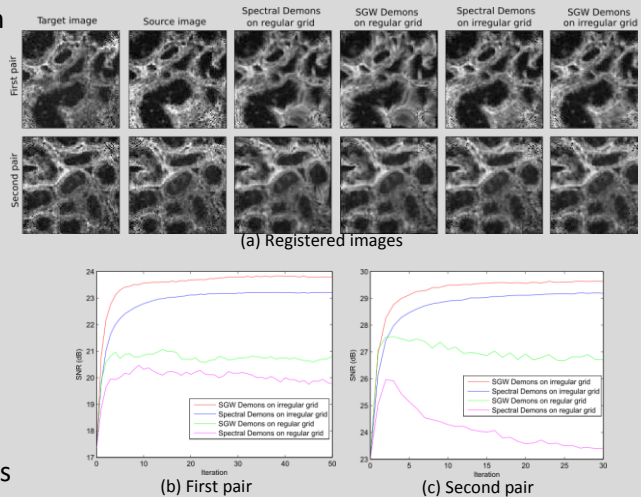


Figure 2: Comparison between the registered images and SNR values of our method and the existing methods for two pairs of images ($\alpha_g = 1$, $\alpha_s = 5$, $\alpha_t = 20$ for registration on irregular grid)

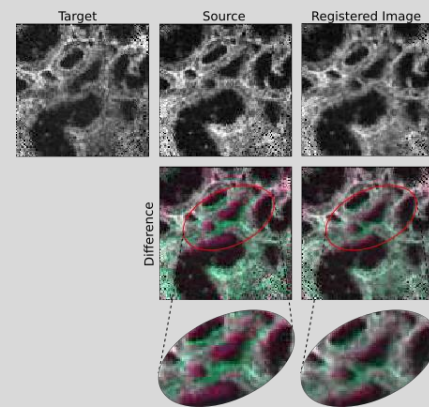


Figure 3: Registration of the first pair on irregular grids (Red channel: target image, Green channel: source image)

Conclusion et perspectives :

Our method shows that the Log-Demons framework can be adapted to the registration of images with floating point coordinates. The experimental results show that this application on the raw coordinates without any projection on Cartesian coordinate system performs better than the classical one on a regular grid. In all tested cases, the registration based on Graph Wavelets is more robust than the Graph Spectral based method. We believe that, our work will motivate other researches on image processing on irregular grids. Our future work will focus on studying the impact of graph construction and to create a complete panorama from the image sequence.

Références (5 max.):

- [1] Vercauteren, T., Pennec, X., Perchant, A., Ayache, N.: Symmetric log-domain diffeomorphic registration: A demons-based approach. Medical Image Computing and Computer-Assisted Intervention MICCAI 2008 pp. 754-761 (2008)
- [2] Hammond, D.K., Vandergheynst, P., Gribonval, R.: Wavelets on graphs via spectral graph theory. Applied and Computational Harmonic Analysis 30(2), 129-150 (2011)
- [3] Lombaert, Herve, et al. "Spectral Demons—image registration via global spectral correspondence." *Computer Vision—ECCV 2012*. Springer, Berlin, Heidelberg, 2012. 30-44.
- [4] Pham, Nhung, et al. "Spectral Graph Wavelet based Nonrigid Image Registration." *2018 25th IEEE International Conference on Image Processing (ICIP)*. IEEE, 2018.
- [5] Ducourthial, Guillaume, et al. "Development of a real-time flexible multiphoton microendoscope for label-free imaging in a live animal." *Scientific reports* 5 (2015): 18303.



Titre :

Classification des signaux EEG pour BCI à base de descripteurs de Fourier et réseaux de neurones

Auteurs :

Kathia CHENANE, Youcef TOUATI, Larbi BOUBCHIR, Boubaker DAACHI
LIASD, Université Paris 8, France

Résumé (100-200 mots) :

Cet article propose une méthode de classification des signaux ÉlectroEncéphaloGraphiques (EEG) pour les Interfaces Cerveau-Machine (BCI : Brain Computer Interface). Une BCI est un système composé d'une architecture matérielle embarquant du logiciel, dédié principalement pour l'interprétation des activités cérébrales permettant ainsi aux personnes handicapées d'interagir de manière directe ou indirecte avec leur environnement immédiat selon des modèles de comportement et de réflexion associés. Dans cette étude, nous présentons un modèle de classification des signaux EEG issus de cinq activités cérébrales différentes acquises par l'utilisation d'une méthode non-invasive. Un algorithme de quantification basée sur la transformée de Fourier suivi d'un modèle d'extraction de caractéristiques pour le calcul de la moyenne quadratique (RMS) sont mis en œuvre pour améliorer la qualité des données mesurées palliant de ce fait la possibilité de perte d'informations et d'artefacts. Afin de garantir une meilleure efficacité du processus de prise de décision, nous proposons une approche neuronale du type Self-Organizing Map (SOM) permettant une meilleure classification des différentes activités traitées.

Mots-clefs (3 à 5 max.) :

BCI, EEG, descripteurs de Fourier, classification, réseaux de neurones

Contexte et état de l'art :

Une BCI peut être vue comme un système permettant à l'utilisateur de communiquer uniquement par le biais de son activité cérébrale correspondant à des signaux électriques, traduite par des actions ou commandes. Dans la littérature, il existe principalement deux types d'acquisition des signaux électriques : une acquisition invasive qui nécessite de disposer directement une matrice d'électrodes sur le cerveau d'un individu et une acquisition non-invasive qui consiste à placer les électrodes tout simplement sur le cuir chevelu. Cette dernière est de loin la plus utilisée pour des raisons pratiques même si elle engendre une perte d'informations due à la présence de bruits de mesure. Ainsi, une exploitation efficace des signaux EEG mesurés revient à mettre en œuvre des procédures adéquates de prétraitement et de filtrage permettant une meilleure interprétation des différentes activités cérébrales [1]. Une fois les bruits et les artefacts réduits, une procédure d'extraction des caractéristiques est appliquée pour alimenter le module de classification dans le but de distinguer les intentions de l'utilisateur en lisant son activité cérébrale.

Parmi les méthodes de classification [2], on peut distinguer l'Analyse Discriminante Linéaire (LDA), la Machine à Vecteur de Support (SVM) ou encore les Réseaux de Neurones Artificiels (RNA) qui, contrairement aux précédents, sont plus adaptés pour le traitement de classes multiples. Dans notre étude, on se basera sur l'approche RNA.

Travail proposé (Décrire l'objectif du travail et indiquer clairement le problème/défi technique étudié) :

Notre travail consiste à proposer un classifieur permettant d'interpréter la pensée d'un individu en fonction des activités cérébrales mesurées à travers des dispositifs type BCI. Sa mise en œuvre intègre trois étapes indispensables à savoir, l'acquisition, le traitement et la classification. Ainsi, la démarche adoptée consiste à effectuer dans un premier temps, une série d'enregistrement de signaux EEG correspondant à des activités

cérébrales de quatre sujets *a priori* sains. Pour ce faire, notre choix s’est porté sur cinq activités différentes : *converser*, *lire*, *regarder une vidéo*, *se reposer* et *travailler*. L’activité ‘*converser*’ s’effectue en Anglais et implique le sujet concerné et une deuxième personne. L’activité ‘*lecture*’ porte sur un texte également écrit en Anglais sans illustration et L’activité ‘*regarder une vidéo*’ correspond à un documentaire sur la nature en Français. Concernant les deux activités ‘*se reposer*’ et ‘*travailler*’, le sujet doit respectivement se détendre et résoudre un problème mathématique. D’un point de vu pratique, nous avons utilisé un dispositif expérimental de type g.GAMMASys comprenant 18 électrodes ou canaux dont une pour la REFERENCE et l’autre pour le GROUND. La durée d’expérimentation liée à l’acquisition des signaux EEG correspondant à la mesure d’une seule activité, est de 120 secondes sur trois cycles étalés à intervalles de temps différents dans la journée. Une base de données constituée de données d’apprentissage est construite avec plus de 30436 enregistrements par canal. La Figure 1 illustre l’emplacement physique des différents capteurs et les zones cibles selon le système international 10-20 [3]. L’acquisition des données d’enregistrement est réalisée avec un pas et une fréquence d’échantillonnage respectivement de 0.0039 secondes et 256Hz.

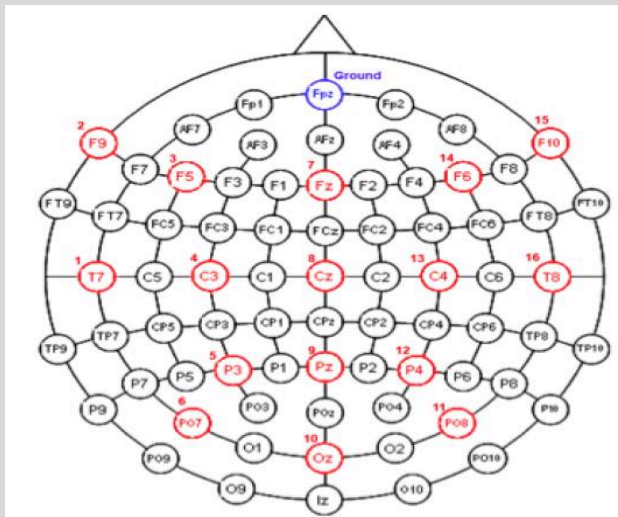


Figure 1 : Emplacement des électrodes. [3]

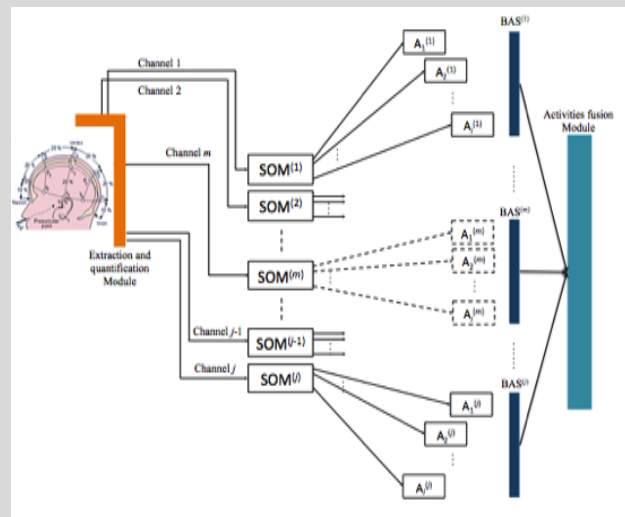


Figure 2 : Implémentation de SOM.

La récupération des signaux utiles passe par l'utilisation d'un filtrage fréquentiel par la transformée de Fourier rapide (FFT : Fast Fourier Transform). Cette dernière est générée par des algorithmes rapides dont la complexité est proportionnelle à $n \log_2 n$ appliqués sur la transformée de Fourier discrète. Les valeurs réelles sont obtenues par l'utilisation de RMS [4].

Une fois obtenus, une procédure de classification est effectuée en implémentant un modèle neuronal type SOM [5]. Ce dernier permet d'effectuer une classification de plus de deux activités avec un apprentissage non supervisé. Sa mise en œuvre passe par deux phases : une phase d'apprentissage aléatoire et cyclique qui commence par une initialisation des 16 vecteurs correspondant aux différents canaux d'acquisition. Ainsi, l'ensemble des neurones dont le statut correspond à des neurones gagnants appelés Best Match Unit (BMU) calculent les distances euclidiennes par rapport à leurs voisinages et en fonction des résultats obtenus et de leur degrés d'appartenance, ils seront associés à une classe ou activité la plus proche. L'approche proposée est illustrée dans la figure 2. Les paramètres A , BAS et j correspondent respectivement à l'activité considérée, l'activité sélectionnée et au nombre d'électrodes (1-16).

Résultats :

Les figures 3 et 4 illustrent les résultats obtenus. Ils correspondent aux différentes activités cérébrales d'un des quatre sujets. Les cinq différentes activités cérébrales obtenues par les réseaux SOM correspondant aux différents capteurs (16 canaux), sont différenciées grâce à des couleurs (*converser*→Rouge, *lire*→Bleu, *se reposer*→Jaune, *regarder une vidéo*→Vert, *travailler*→Rose).

Les résultats obtenus montrent l'efficacité de l'approche proposée tout au long de la chaîne de traitement des données. A titre d'exemple, l'activité '*converser*' fait intervenir l'ensemble des parties du cerveau avec des degrés différents plus ou moins important, confirmant les résultats illustrés ci-dessous puisque la couleur

Rouge correspondante apparaît dans l'ensemble des réseaux SOM. L'activité 'regarder une vidéo' fait intervenir le cortex visuel, ici représentée par la couleur Vert. En effet, l'électrode 16 disposée dans la partie occipitale comme l'illustre la figure 1, atteste de cela. La partie la plus prépondérante dans la figure 4 pour le canal 16 est représentée par la couleur Vert.

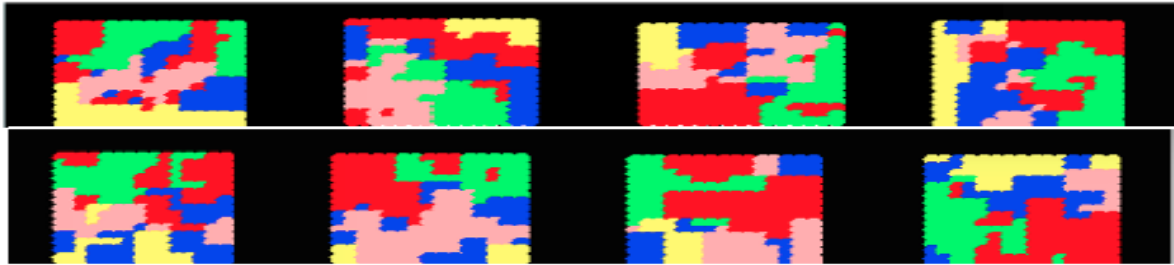


Figure 3 : Résultats des électrodes 1 à 8 obtenus après implémentation de SOM.

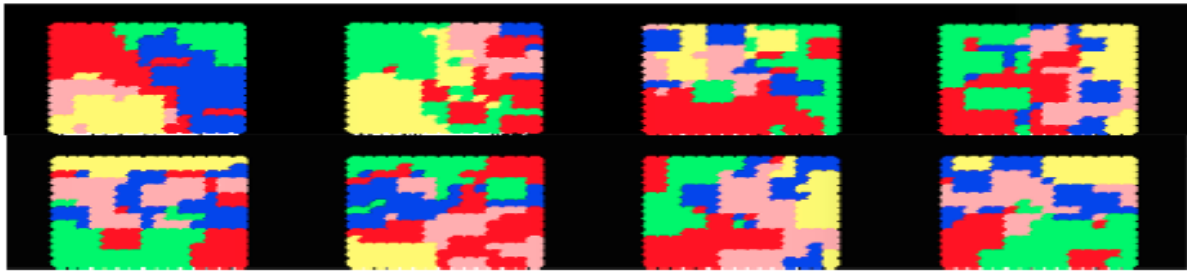


Figure 4 : Résultats des électrodes 9 à 16 obtenus après implémentation de SOM.

Conclusion et perspectives :

Dans cette étude, nous avons développé une approche basée sur l'utilisation des réseaux SOM pour la classification des signaux EEG pour les BCI. Pour ce faire, nous avons présenté un modèle de classification issu de cinq activités cérébrales différentes acquises à travers l'utilisation d'une méthode non-invasive. En outre, nous avons mis en œuvre un algorithme de quantification basée sur la transformée de Fourier suivi d'un modèle d'extraction de caractéristiques pour le calcul de la moyenne quadratique afin d'améliorer la qualité des données mesurées et réduire les artefacts. Les résultats obtenus sont intéressants et prometteurs. Toutefois, des améliorations doivent être apportées sur la partie classification pour une meilleure prise de décision particulièrement d'un point de vu temps réel. L'idée est d'utiliser la pensée de l'utilisateur en lui associant une interprétation visuelle (*neurofeedback*).

Références (5 max.):

- [1] Lotte, F., Bougrain, L., Cichocki, A., Clerc, M., Congedo, M., Rakotomamonjy, A., & Yger, F.A "review of classification algorithms for EEG-based brain-computer interfaces: a 10 year update", Journal of neural engineering, 2018, vol. 15, no 3, p. 031005.
- [2] Bharne, P. P., & Kapgate, D.A "Review of Classification Techniques in Brain Computer Interface", International Journal of Computer Sciences and Engineering, 2014, vol. 2, p. 68-72.
- [3] Jasper, H. "Report of the committee on methods of clinical examination in electroencephalography". Electroencephalography Clinical Neurophysiology, vol. 10, no. 2, p. 370-375, 1958.
- [4] Hindarto, H., & Sumarno, S. "Feature extraction of electroencephalography signals using fast fourier transform", CommIT (Communication and Information Technology) Journal, 2016, vol. 10, no 2, p. 49-52.
- [5] Kohonen, T. "The self-organizing map", Proceedings of the IEEE, vol. 78, no. 9, pp. 1464--1480, 1990.



Titre :

Compression d'images couleur par classification adaptative d'histogrammes en fonction de la densité informationnelle.

Auteurs :

CLÉMENT Alain, BENOIT Landry

Laboratoire Angevin de Recherche en Ingénierie des Systèmes (LARIS EA7315), Université d'Angers, France.

Résumé (100-200 mots) :

Pour conserver la qualité perceptuelle des images couleur lors de la compression, il est essentiel de préserver l'organisation des couleurs dans les zones de l'image les plus porteuses d'information. Nous proposons une méthode de compression par classification adaptative d'histogramme couleur permettant de réduire efficacement le nombre de couleurs de l'image sans dégrader son apparence visuelle. La méthode exploite une décomposition hiérarchique de l'histogramme en fonction de sa densité d'information. La classification est non supervisée et non paramétrique et conserve la morphologie de l'histogramme couleur. Nous comparons nos résultats à ceux de deux approches non supervisées de référence, la compression JPEG et la compression linéaire par réduction du nombre de bits.

Mots-clefs (3 à 5 max.) :

image couleur, quantification vectorielle, histogramme, densité informationnelle

Contexte et état de l'art :

La compression des images couleur repose sur une triple redondance d'information, spatiale, statistique et psychovisuelle. Les meilleurs taux de compression sont obtenus avec des méthodes destructrices dont la compression est irréversible ; la dégradation perceptuelle de l'image est généralement proportionnelle au taux de compression. De manière intuitive, une compression linéaire simple consiste à réduire le nombre de bits qui codent la dynamique de l'image (quantification scalaire uniforme). Cette approche conduit à une rapide dégradation de l'image. La transformation des vecteurs de base préalablement à la quantification scalaire permet de réduire significativement la dégradation perceptuelle ; c'est le principe utilisé par la très populaire compression JPEG [1]. La quantification vectorielle [2] consiste à effectuer une classification de l'espace de codage des couleurs en tenant compte des corrélations entre les composantes colorimétriques ce que ne fait pas la quantification scalaire. Dernièrement, différentes méthodes de compression vectorielle exploitant la redondance statistique ont été proposées. Dans [3], Yang et al. estiment les variations de la densité informationnelle de l'image par des sommes de lois normales ou polynomiales. Dans [4], Shi et al. utilisent les réseaux de neurones pour classer les images en fonction des variations de densité de leur histogramme couleur. Ces méthodes supposent une connaissance a priori de l'image ou une modélisation paramétrique de l'histogramme couleur. L'histogramme multidimensionnel d'une image couleur est diffus, c'est à dire constitué de points couleur non connexes au sens strict de la connexité. Ces éléments ne forment pas un nuage de points homogène mais se répartissent en zones de densité variable. L'algorithme d'histogramme compact [5] facilite le traitement vectoriel des histogrammes multidimensionnels sans perte d'information.

Travail proposé (*Décrire l'objectif du travail et indiquer clairement le problème/défi technique étudié*) :

Nous proposons une méthode de compression par quantification vectorielle adaptative de d'histogramme multidimensionnel. La méthode exploite une décomposition hiérarchique des modes de l'histogramme en fonction des variations locales de la densité d'information. L'algorithme de classification est non supervisé et non paramétrique ; il conserve la morphologie de l'histogramme couleur en gardant inchangée la colorimétrie des zones les plus porteuses d'information.

L'histogramme multidimensionnel d'une image couleur est la représentation d'une fonction discrète ν qui à chaque couleur $\vec{X}$ de l'image associe le nombre de pixels ayant cette valeur (Eq. 1). Il peut être vu comme la densité de probabilité d'une variable aléatoire dont l'image constitue un ensemble de réalisations.

$$\nu : \mathbb{R}^3 \rightarrow \mathbb{N}, \vec{X} \mapsto \nu(\vec{X})$$

(1)

Issue de la théorie de l'information, l'entropie de Shannon mesure la quantité d'information portée par un groupe. Elle se définit comme : $E(I) = - \sum_{i=1}^N p_i \cdot \log_2(p_i)$ avec $p_i = v(\vec{X}_i) / N$ (2)

où N est le nombre de pixels de l'image I et p_i la probabilité a posteriori donnée par l'histogramme. Pour une partition en K classes de l'histogramme, l'équation 2 se réécrit :

$$E(I) = \sum_{j=1}^K E(C_j) \text{ et } E(C_j) = - \sum_{k \in D_j} p_k \cdot \log_2(p_k) \text{ avec } D_j \text{ le domaine de la classe } C_j \quad (3)$$

L'entropie est indépendante de la taille de l'image. Elle est un indicateur de la complexité colorimétrique et permet de comparer entre eux des histogrammes. Elle peut donc servir à mesurer la performance d'une compression par quantification du nombre de couleurs. Calculée sur un voisinage local de l'histogramme, l'entropie tient compte à la fois du nombre de pixels et du nombre de couleurs de l'image pour ce voisinage ; elle mesure alors la densité informationnelle de l'histogramme. L'histogramme d'une image couleur (Fig.1) présente des zones de densité informationnelle variable en raison de la non-uniformité de la répartition des points couleur. L'histogramme est diffus, c'est à dire constitué de points couleur non connexes au sens strict de la connexité. Nous définissons l'étiquetage en composantes connexes diffuses (ECCD) d'un histogramme H à P dimensions par extension de la connexité classique sur un voisinage V P -dimensionnel défini par la distance de Tchebychev d (norme L^∞) :

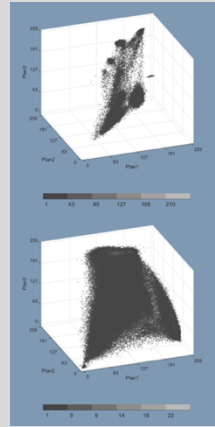
$$V(\vec{X}_i) = \cup_d V_d(\vec{X}_i) \text{ où } V_d(\vec{X}_i) = \{\vec{X}_j \in H, \text{dist}(\vec{X}_i, \vec{X}_j) = d\} \text{ et } \text{dist}(x,y) = \max_{i=1}^n |x_i - y_i| \quad (4)$$

Une partition de l'histogramme par ECCD a pour avantage de conserver la morphologie de l'histogramme. Pour prendre en compte les variations de densité informationnelle, nous faisons varier d en construisant une arborescence des composantes connexes diffuses par décomposition récursive des modes de l'histogramme. Le critère d'arrêt de la décomposition de chaque composante connexe est une valeur minimum d'entropie en % de l'entropie totale de l'image (Eq. 3). Les composantes dont l'entropie est inférieure au seuil sont éliminées. L'algorithme est initialisé avec d tel que l'histogramme ne forme au départ qu'une seule composante connexe. d décroît ensuite récursivement jusqu'à un minimum puis la décomposition des modes se poursuit par seuillage d'effectifs croissants comme dans un algorithme de watershed. La décomposition est non supervisée et ne fait aucune hypothèse statistique sur l'organisation de l'histogramme. La classification des valeurs colorimétriques non retenues lors de l'étiquetage des composantes connexes est ensuite effectuée en attribuant à chaque composante la valeur de son centre de masse. La classification respecte la contrainte hiérarchique induite par l'arborescence des classes en affectant tout d'abord un point couleur à sa classe parent avant de rechercher parmi les descendants la classe fille la plus proche. La métrique de Tchebychev (Eq. 4) est utilisée. L'algorithme de quantification vectorielle a été développé à P -dimensions sous Matlab®. Il utilise le codage d'histogramme compact [5] qui permet le traitement vectoriel d'histogrammes de grande taille sans perte d'information. Comme l'histogramme compact n'est pas une structure linéaire, les algorithmes classiques d'étiquetage ne sont pas applicables. Un algorithme d'étiquetage en composantes connexes diffuses (ECCD) pour histogramme compact a été spécifiquement développé en langage C et compilé pour Matlab® (Mex-file).

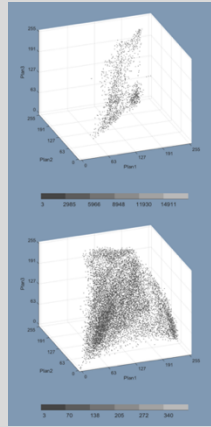
Résultats :

Nous présentons nos résultats sur des images RVB 24 bits de la base USC-SIPI (<http://sipi.usc.edu>). Le taux de compression TC est défini comme le rapport dans l'histogramme entre le nombre de points couleur après classification et le nombre de points couleur d'origine. Ce taux dépend du minimum d'entropie des classes (critère d'arrêt de l'algorithme). La figure 1 montre les résultats de quantification vectorielle pour trois ordres de grandeur de TC . Les valeurs numériques sont précisées dans le tableau 1. On constatera la conservation de la morphologie de l'histogramme même quand la compression est forte. Dans les figures d'histogrammes, les plans 1, 2 et 3 correspondent aux axes colorimétriques RVB. Le 4ème axe est constitué par les effectifs des cellules non vides dont la valeur est donnée par la barre de niveaux de gris. Nous comparons dans le tableau 1 les images originales, la compression JPEG (format standard, qualité moyenne), la quantification scalaire (QS) par réduction du nombre de bits et la quantification vectorielle (QV). La complexité colorimétrique est évaluée par trois grandeurs : le nombre de couleurs (nombre de cellules non vides de l'histogramme), le nombre moyen de pixels par point couleur et la mesure d'entropie de Shannon. Le nombre de bits nécessaire au codage de l'image est précisé. Ces résultats confirment que la réduction du nombre de couleurs est bien plus forte avec la quantification scalaire ou vectorielle qu'avec la compression JPEG dont les valeurs d'entropie restent élevées. Contrairement à la compression scalaire, notre compression vectorielle préserve la qualité perceptuelle des images (Fig. 2) et ne fait pas apparaître d'effet de bloc comme la compression JPEG.

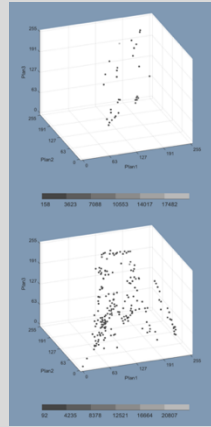
images et histogrammes non compressés



TC = 95%



TC = 99,9%



TC = 99,99%

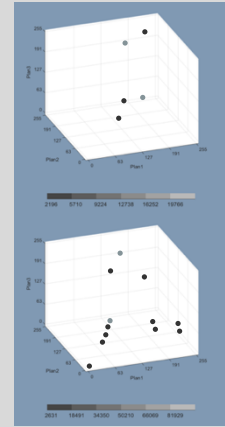


Figure 1 : quantification vectorielle des histogrammes couleur

		original	JPEG	QS			QV		
House 256x256	TC	0	41,43	94,93	99,95	99,98	96,06	99,90	99,99
	points couleur	33925	19869	1721	18	8	1336	33	5
	entropie	13,8	11,9	7,0	2,3	1,8	7,7	3,9	2,0
	pixels/point couleur	1,9	3,3	38	3641	8192	49	1986	13107
	bits	24	24	15	6	3	11	6	3
Mandrill 512x512	TC	0	32,09	96,50	99,91	99,99	95,28	99,91	99,99
	points couleur	230427	156486	8054	209	8	10869	199	12
	entropie	17,7	17,0	11,6	6,1	2,5	12,7	6,9	2,8
	pixels/point couleur	1,1	1,7	33	1254	32768	24	1317	21845
	bits	24	24	15	9	3	14	8	4

Tableau 1 : comparaison des méthodes de compression

QV



QS



Figure 2 : images de quantification vectorielle (QV) et de quantification scalaire (QS) pour TC = 99,99%

Conclusion et perspectives :

L'objectif de ce travail était de proposer une méthode de compression qui réduise efficacement le nombre de couleurs de l'image sans dégrader son apparence visuelle. Notre algorithme de quantification vectorielle effectue une classification de l'histogramme qui s'adapte aux variations locales de la densité informationnelle. Il est non supervisé et conserve la morphologie de l'histogramme couleur. Nous montrons que notre méthode réduit plus efficacement le nombre de couleurs que la compression JPEG sans dégrader l'image comme la quantification scalaire. Notre méthode est indépendante du nombre de composantes de l'image et permet le traitement d'histogrammes de grande taille sans perte d'information. La méthode proposée convient à tout type d'image couleur et plus généralement multicomposante.

Références (5 max.):

- [1] W. Pennebaker, J. Mitchell, *JPEG Still Image Data Compression Standard*, Springer, New York, 1993.
- [2] C. Charrier, H. Cherifi, *Fast vector quantization algorithm preserving color image quality*, proceedings of the SPIE, 1998, vol. 3308, pp. 93-104.
- [3] A. Y. Yang, J. Wright, Y. Ma, S. S. Sastry, *Unsupervised segmentation of natural images via lossy data compression*, Computer Vision and Image Understanding, 2008, vol. 110, pp. 212-225.
- [4] Y. Shi, Z. Chen, Z. Qi, F. Meng, L. Cui, *A novel clustering-based image segmentation via density peaks algorithm with mid-level feature*, Neural Computing and Applications, 2017, vol. 28 suppl. 1, pp. 29-39.
- [5] A. Clément, L. Busin, O. Lezoray, L. Macaire, *Region Segmentation*, eds. C. Fernandez-Maloigne, F. Robert-Inacio, L. Macaire, Digital Color Imaging, Wiley-ISTE, Digital Signal and Image Processing Series, London, 2012, pp. 149-191.



Titre:

Design of optimal wavelet image compression scheme based on spline transform and improved coding algorithm

Auteurs :

Rania Boujelbene¹, Yousra Ben Jemaa¹, Larbi Boubchir²

¹U2S Laboratory, Tunis-El Manar, Tunis, Tunisia

²LIASD research Lab., University of Paris 8, France

Résumé (100-200 mots):

In this paper, a new wavelet image compression scheme is proposed, which is based on an optimal spline wavelet transform and an improved Embedded zero-tree wavelet algorithm. A comparative study of the proposed approach with the existing works is carried out. The obtained results demonstrate the efficiency of the proposed approach to achieve significant gains in terms of Peak Signal to Noise Ratio (PSNR) and structural similarity index measure (SSIM) for reconstructed image.

Mots-clefs (3 à 5 max.):

Image compression, spline wavelet transform, coding algorithm, EZW.

Contexte et état de l'art :

Among the existing lossy compression techniques, transform coding is considered as one of the most efficient strategies. The basic components of this technique are carried out in three steps, namely the image transformation step tracked by the quantization and the entropy coding steps. To the best of our knowledge, most existing works have mainly focused on both transformation and coding steps. For instance, in [1], the authors focused on the transformation step. In fact, they present a new biorthogonal wavelet transforms using polynomial splines. They study the influence of the filters order on lossy image compression in order to choose the best transform. In [2], the authors check the success of the different biorthogonal spline wavelet transforms and justify the choice of the optimal transform presented in [1]. In [3], the authors proposed an extensive study of the different existing wavelets and the encoding algorithms. Recently, in [4], the authors propose an extensive study of the recent improvements of the various coding algorithms. They compare and analyze the performance of the recent enhancement and compression techniques. Indeed, this study allows to conclude that the EZW algorithm is more promising than Set Partitioning in Hierarchical Trees (SPIHT) algorithm and can bring more improvements. Unlike [1-4], we dealt with the two steps of image compression in order to design a new efficient wavelet image codec. Our contribution is emphasized as follows:

- We propose an efficient codec for lossy image compression.
- We illustrate the efficiency of the proposed codec through a comparative study with the existing works in terms of PSNR and SSSIM.

Travail proposé(Décrire l'objectif du travail et indiquer clairement le problème/défi technique étudié) :

The block diagram for our optimal codec is shown in Figure 1. It is composed of two connected blocs, namely spline wavelet transform and improved EZW algorithm. The spline wavelet transform is performed in a lifting manner. This scheme minimizes the number of operations as well as the memory occupation. In addition, it is a refined system of very short filters reaching significant computational and memory savings.

The polynomial splines is used as a tool for the construction of the prediction and update filters. Indeed, they provide filters that have properties valuable for image processing. The impulse responses of these filters are

symmetric and well localized in spatial domain. Moreover, they have flat spectra and they generate wavelets that can have any number of vanishing moments. Specifically, the biorthogonal wavelet transform generated by polynomial spline of fifth order has been selected. That has been tied to the fact that it has good properties and achieves good performances in terms of PSNR and encoding time as demonstrated in [1,2]. Furthermore, it outperforms the biorthogonal B9/7 transform [5] which is frequently used in image compression.

After the wavelet transform step, we try to get a way to code the wavelet coefficients into an effective result by taking into account the storage space and the redundancy. So, from the study done on the different wavelet algorithms presented in [4], we conclude that the Embedded zero-tree wavelet algorithm (EZW) algorithm is more promising than the other coding schemes and can bring more improvements.

Therefore, another improved version of the EZW algorithm is proposed which combines the two approaches presented in [6,7] and exploits their limitations. The main goal of the proposed method is to increase the number of coefficients not to be encoded and to reduce the total number of symbols. In addition, the coding bits is decreased by the use of a compressor operator. Hence, this allows to achieve a high compression performance in terms of PSNR and bitrate.

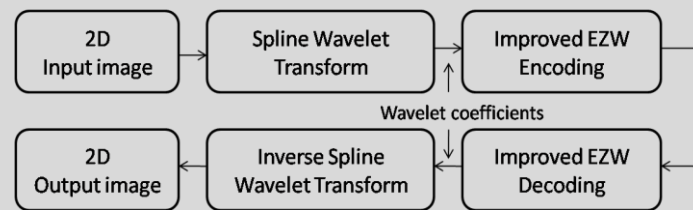


Figure 1.Block diagram for the proposed codec

Résultats :

To assess the efficiency of the proposed approach, numerical experiments have been performed on a number of test images. In this paper, we present just the results of the two test images. PSNR and SSIM were used to evaluate the quality and the distortion of the reconstructed image.

Images	Bitrate(bpp)	Proposed codec		Previous work[2]		JPEG2000	
		PSNR(dB)	SSIM	PSNR(dB)	SSIM	PSNR(dB)	SSIM
Lena	0.25	34.57	0.9555	34.08	0.9549	34	0.954
	0.5	37.53	0.9779	37.2	0.9775	37.3	0.9777
	1	40.88	0.99	40.02	0.9897	40.4	0.989
Barbara	0.25	28.21	0.901	27.5	0.8933	27.5	0.894
	0.5	33.1	0.9574	31.69	0.9542	32.6	0.9561
	1	38.04	0.9901	36.62	0.9848	37.45	0.987

Table1.Lossy performance comparison of different image codecs for several grayscale test images

Table 1 shows the results provided by the three codecs, and assessed at the same bitrates. According to this Table, it can be clearly seen that the proposed codec outperforms the other codecs for all bit-rates on the test images. At the same compression ratio, the PSNR values of the reconstructed images with the proposed technique are higher than those for the other existing works. Clearly, the improvement over the standard JPEG2000 is significantly high around 0.23-0.71 dB and up to 1.4 dB over the work proposed in [2]. Also, it is observed from Table 1 that the proposed method produces high SSIM values when compared to the other methods.

We display in Figure 2 a zoom from the reconstructed Barbara image using the three codecs for RC=0.25bpp. We can see that the texture of the tablecloth and the scarf is better disclosed with our proposed codec.



(a)

(b)

(c)

Figure 2. Fragment of Barbara reconstructed using: a) proposed codec, b) previous work [2], c) JPEG2000 for RC= 0.25 bpp

Conclusion et perspectives :

In this paper, an efficient codec for image compression has been proposed. It is based on an optimal polynomial spline wavelet transform of order 5 and on an enhanced EZW algorithm. Experimental results have shown the superiority of the proposed method over existing ones in terms of PSNR and SSIM, for different test images. For our future work, we will try to further improve the performance of the proposed codec by proposing another enhancement to achieve a high compression gain in terms of different performance criteria.

Références:

- [1] R. Boujelbene, Y. B. Jemaa and M. Zribi, "Toward an optimal b-spline wavelet transform for image compression", In International Conference on Computer Systems and Applications, pp. 1-8, Morocco, 2016.
- [2] R. Boujelbene, Y. B. Jemaa and M. Zribi, "An efficient codec for image compression based on spline wavelet transform and improved SPIHT algorithm", In International Conference on High Performance Computing & Simulation , pp. 819-825, Italy, 2017.
- [3] U. Rasool, S. Mairaj, T. Nazeer, and S. Ahmed, "Wavelet based image compression techniques: Comparative analysis and performance evaluation", International Journal of Emerging Technologies in Engineering Research, 2017.
- [4] R. Boujelbene, Y. B. Jemaa and M. Zribi, "A comparative study of recent improvements in wavelet-based image coding schemes", Multimedia Tools and Applications, pp. 1-35, 2018.
- [5] A. Cohen, I. Daubechies, and J.-C. Feauveau, "Biorthogonal bases of compactly supported wavelets", Communications on pure and applied mathematics, pp. 485-560, 1992.
- [6] T. Brahimi, F. Laouir, L. Boubchir and A. Ali-Chérif, "An improved wavelet-based image coder for embedded greyscale and colour image compression", AEU-International Journal of Electronics and Communications, vol. 73, pp. 183-192, 2017.
- [7] S. Che, Z. Che, H. Wang and Q. Huang, "Image compression algorithm based on decreasing bits coding", In Fifth International Conference on Information Assurance and Security, pp. 217–220, China, 2009.



Titre :

Feature extraction for face recognition based on SSA-NIG and statistical features

Auteurs :

Basma Ammour¹, Larbi Boubchir², Toufik Bouden¹, Messaoud Ramdani³

¹NDT Laboratory, University of Jijel, Algeria

²LIASD research Lab, University of Paris 8, France

³LASA research Lab, University of Badji Mokhtar-Annaba, Algeria

Résumé (100-200 mots) :

Face recognition system is based much more on feature extraction techniques. In order to capture pertinent features and to overcome difficulties faced by pose variation, illumination variation, and expression of face recognition, many feature extraction techniques have been proposed in the literature, such as Local Binary Pattern (LBP) and its versions, modified Zernike moments, histogram oriented gradient (HOG), etc. This paper proposes a new feature extraction technique combining Singular Spectrum Analysis (SSA) modeled with Normal Inverse Gaussian (NIG) model and statistical features (entropy, energy and skewness) derived from wavelet transform. The proposed method is used for the implementation of face identification system and the classification process can be done by two classifiers K-Nearest Neighbor (KNN) and Support Vector Machines (SVM). Experimental results carried out on FERET and ORL face databases. The proposed face identification system achieved good recognition rate.

Mots-clefs (3 à 5 max.) :

Face recognition, feature extraction, SSA, NIG, classification.

Contexte et état de l'art :

There is an increasing demand for accurate and secure face biometric system in different applications and places, such as in mobile phone, in laptops, in airports and so on [1]. This biometric has many advantages against other biometrics like iris and fingerprint modalities, is the natural way to recognize a person, non-intrusive modality, captured from a distance, etc. Several face recognition systems have been proposed in the literature, hence there is no reliable face recognition system, which is capable to solve all the related problems. Among these proposed systems, principal component analysis (PCA) and linear discriminant analysis (LDA) are the most known global feature extraction in image processing, there is also locality preserving projection (LPP), unsupervised discriminant projection (UDP), local discriminant embedding (LDE), and so on. But local features are more robust to deal with local changes such as illumination, scale, pose and so on. Gabor filter and LBP are two popular methods for texture feature extraction. They are based also an hybrid feature extraction which combined local features and global features in order to investigate the advantages of each other and generates robust features [2].

Travail proposé (Décrire l'objectif du travail et indiquer clairement le problème/défi technique étudié) :

This paper proposes a new feature extraction method for face recognition based on statistical features generated from SSA-NIG and wavelet transform in order to extract relevant information and improve the recognition rate. The face recognition system contains three main phases: pre-processing, feature extraction and classification, described as follows:

1. Face image is enhanced then Viola-Jones algorithm is used to detect the face image and facial features (such as nose, eyes and mouth) as shown in the Fig. 1.
2. Feature extraction performed with SSA-NIG and statistical features of wavelet, SSA is mostly used in time series analysis [3] it decomposes the signal into a sum of principal components (PC) as shown in Fig. 2. In facial image feature extraction, 1D SSA uses the pixels of images as a signal, then the signal is decomposed in PC, in order to extract pertinent information and eliminate noisy data caused with illumination variation and expression variation. NIG pdf is applied for modeling the chosen PC component of SSA signals generated from the detected face, nose, eyes and mouth sub-images. The effect of mean and variance on the shape of NIG pdf's are shown in Fig. 3. The NIG pdf is a variance-mean mixture density function, in which the mixing distribution is the inverse Gaussian density [4], it models the histogram of the SSA signal. Fig. 4 shows the corresponding NIG pdfs constructed from the estimated mean and variance (in red). On the other hand,

statistical features (energy, entropy and skewness) of wavelet sub-band are also used. Daubchies wavelet transform is used to perform decomposition at level 2. In this level, we calculating the energy, entropy and skewness at this level.

3. The classification process is performed with KNN and SVM, in order to improve the recognition of the system and enhance their performances by choosing the best classifier that gives the best recognition rate.

Fig 1. diagram of face detection and pre-processing.

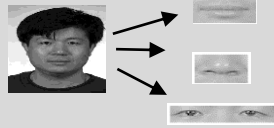


Fig 2. 1D-SSA of facial image signal with M=6

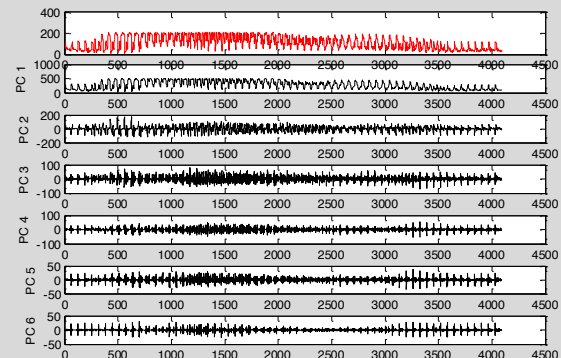


Fig 3. Effect of α and δ on the shape of NIG pdf.

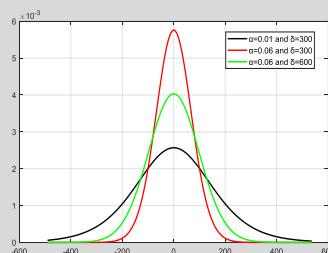
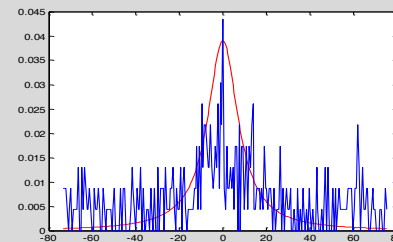


Fig 4. Corresponding NIG pdfs constructed from the estimated mean and variance (in red)



Résultats :

For the evaluation, two databases are used: ORL and FERET face. The recognition rate is given for different components PC and width M of SSA algorithm. The process of evaluation is detailed in [5].

Results on ORL database:

	SVM				
	PC1	PC2	PC3	PC1+ PC2	PC1+ PC2+PC3
M= 8	99.00%	94.50%	90.25%	97.75%	98.00%
M= 4	98.25%	83.50%	79.75%	93.66%	91.25%
	KNN				
	PC1	PC2	PC3	PC1+ PC2	PC1+ PC2+PC3
M= 8	96.50%	96.00%	89.00%	96.5%	95.5%
M= 4	96.00%	91.50%	86.5%	91.75%	91.00%

Note that the best recognition rates are 99.00% and 98.25% obtained by using the first PC1 and the width of M=8 for the SVM and KNN classifiers, respectively.

Results on FERET database:

	SVM				
	PC1	PC2	PC3	PC1+ PC2	PC1+ PC2+PC3
M=8	93.5%	86.00%	78.00%	93.5%	90.5%
M=4	89.00%	71.33%	70.00%	84.50%	77.00%
	KNN				
	PC1	PC2	PC3	PC1+ PC2	PC1+ PC2+PC3
M= 8	92.00%	84.5%	78.33%	92.00%	90.00%
M= 4	89.5%	78.00%	66.75%	85.25%	75.66%

From this table, the best recognition rates are 93.50% and 92.00% obtained by used the first PC1 and the width of M=8 for the SVM and KNN classifiers respectively.

Experimental results demonstrate the robustness of the proposed feature extraction method, which is efficient against expression variation like in ORL database and a little against illumination variation in FERET database.

Conclusion et perspectives :

The proposed face recognition system is based on SSA-NIG, local and global statistical facial features, and wavelet transform. The proposed feature vector investigates the advantages of local and global features to extract robust features. It gives the best recognition rates up to 99.00% on ORL database and up to 93.5 % on FERET

database. As perspective we can used other modality like iris and fingerprint combined with the face like multimodal system to enhance the security of the biometric system.

Références (5 max.):

- [1] J. Qian, J. Yang, Y. Xu, "Local Structure-Based Image Decomposition for Feature Extraction with Applications to Face Recognition", IEEE Transactions on Image Processing, vol. 22(9), pp. 3591-3603, 2013.
- [2] V. V. Pali, S. Goswami, L. P. Bhaiya, "An Extensive Survey on Feature Extraction Techniques for Facial Image Processing", International Conference on Computational Intelligence and Communication Networks, 2014.
- [3] H. Hassani, Singular Spectrum Analysis: Methodology and Comparison, Journal of Data Science 5 (2007), 239-257
- [4] A. R. Hassan, M. I. H. Bhuiyan, "An automated method for sleep staging from EEG signals using normal inverse Gaussian parameters and adaptive boosting", Neurocomputing, vol. 219, pp. 76-87, 2017.
- [5] B. Ammour, T. Bouden, L. Boubchir, S. Biad, "Face Identification using Local and Global Features", The 40th International Conference on Telecommunications and Signal Processing (TSP), 2017.



Title :

TREE DENSITY ESTIMATOR USING SCALE-SPACE BLOBS AND SADDLE FEATURES.

Autors :

Abril Fernanda García Ramírez, Benoit Tremblais, Christine Fernandez-Maloigne, Matthias Rös.

Abstract:

Climate change is a reality from which we cannot escape. There has not been adequate monitoring of forest inventories worldwide. It is important to have fast, precise and performant forest monitoring methods. Nowadays, the used methods are very limited in their spatial extension and very difficult to extrapolate into other areas. On the other hand, satellite images generally provide poor detailed information about forest structure. A promising and unexplored alternative is the use of unmanned aerial vehicles (UAV) equipped with visual sensors. The UAV possess the potential to generate forest complexity relevant information onto necessary spatial extensions, unlike the terrestrial point monitoring. This work aims to show the potentiality of feature detection methods using the scale-space paradigm applied to multispectral imagery from UAV in an interdisciplinary work. An innovative method to detect trees in ecological landscape is developed using NDVI images from Red and Near Infrared bands acquired by UAV. The strategy consists to perform a tree density estimator, detecting blobs and saddle points as features from scale-space image.

Keywords :

Scale-Space, Blob/Saddle Feature Detection, UAVs, Multispectral Imagery.

Contexte et état de l'art :

Currently OBIA (Object Based Image Analysis) techniques [1] have been of great help for ecological applications, so it have been merged with the old Geographic Information Systems (GIS) giving rise to GEOBIA (Geographic Object Based Image Analysis). Given the need to handle better resolutions, practical solutions have been sought to obtain images with higher resolution and simpler to process than hyperspectral (satellite) images. A growing interest in innovating remote sensing technologies emerges as the use of UAVs for obtaining aerial images of ecological landscapes (forests, agricultural fields, etc.) [3], [5], [6], [7]. Its attractiveness relies in many aspects such as adaptability to imagery acquisition (height flight), low costs of equipment, and the very high resolution of images. The resolution of most extended as aerial images range between one meter resolution, [2], [3], satellite data (2-30m) and UAVs imagery resolution varies through 5-15cm depending on altitude of flight (normally low altitudes). In [2] reader can be referred to a review of algorithms on detecting trees or delineating it some of them are LM (local maxima) detection [4], watershed segmentation, region growing, among others. The Scale Space paradigm has been largely studied by scientific community of computer vision [8]. This scheme formalizes the way of representation and extraction of the information contained in an image. Scale-space representation paradigm allows to obtain the complete topological structure of the image. For the problem of detecting tree crowns in forest images, the previous scheme seems attractive to be implemented to aerial images. There are some works such as Hay [9] who have developed a tree detector using the space-scale paradigm. In this work is also added the use of a VI (Vegetation Index) to estimate tree locations on this map of multispectral images acquired by the UAV. The NDVI of an area containing a canopy of dense vegetation will tend to positive values (from 0.3 to 0.8) while the clouds and snow fields will be characterized by negative values of this index. Other Earth targets visible from space include: free water (for example, oceans, seas, lakes and rivers) that have a fairly low reflectance in both spectral bands (at least far from the edges) and, therefore, give very low values of positive or even slightly negative NDVI, soils that generally exhibit near infra-red spectral reflectance somewhat greater than red, and therefore tend to also generate rather small positive NDVI values (between 0.1 to 0.2).

Proposed work :

This work aims to describe an innovative method for tree crown detection based on scale-space paradigm. The proposed density estimator is then calculated using the NDVI (Normalized Difference Vegetation Index) image that highlights vegetation of the rest of the ground and other components of ecological landscape such as water. The image of the vegetation index uses two registered spectral components, red and near infra-red (NIR) since it has been shown that these two components are in which the frequencies in which the vegetation is found are detected with greater precision. The NDVI is calculated from these individual measurements as follows:

$$NDVI = \frac{NIR - RED}{NIR + RED}$$

The maximum local values (blobs) are considered as the highest points of trees and have brightest contrast in image, so that facilitate detection. Because the NDVI image is in gray scale and the clearest values are those

associated with the vegetation, two stages of preprocessing are operated on the image. First a binarization to the image is applied, establishing a threshold of 0.4 and 0.5, this value is justified because the NDVI image brings negative values (from -0.1 to -0.2) as explained in previous section (for water, snow, etc.). Our image range from 0 to 1, taking into account that vegetation values range between 0.2-0.4 of original NDVI then is necessary to move threshold value adding 0.2 for our tests. Secondly, a morphological operation of *closing* (dilation, erosion) is applied to the binarized image.

Blob-Saddle-feature detection at a fixed scale.

Now we proceed to apply the tree detector. The Gaussian kernel is used to blur the image $I(x, y)$ than in our case is the binarization resulting from NDVI image. By using a scale parameter $t = \sigma^2$ the bi-dimensional kernel is defined as:

$$G_t = \frac{1}{2\pi} e^{-\frac{x^2 + y^2}{2t}}$$

Hence, the space-scale $L(x, y; t)$ is defined as:

$$L_t = G_t(x, y; t) * I(x, y)$$

For each point of image I located at the pair $(x, y; t)$ at a fixed scale t the *blobs* can be defined as the positive local maxima (or negative local minima) in space of the normalized Laplacian of image:

$$(x, y; t) = \underset{x, y, t}{\operatorname{arglmax}} \{ t^\gamma |(L_{t,xx} + L_{t,yy})| \}$$

where $\operatorname{arglmax}$ are the arguments for which a local maximum is found using some parameter γ of normalization with $\gamma = 1$. Notice that $L_{t,xx}$ and $L_{t,yy}$ can be calculated as the derivatives of scale space L_t in the horizontal and vertical directions respectively. The values of saddle points at a fixed scale t can be defined as follows:

$$\nabla L(x, y; t) = (L_x, L_y, L_t)^T = 0$$

The maximum values, known as blobs will be detected using a method developed by Blom [10] called hexagonal grid. In this method, intensity of pixel analyzed is compared with neighboring pixels around it as is depicted in figure 1:

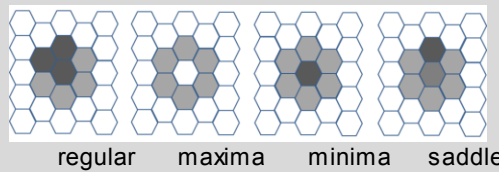
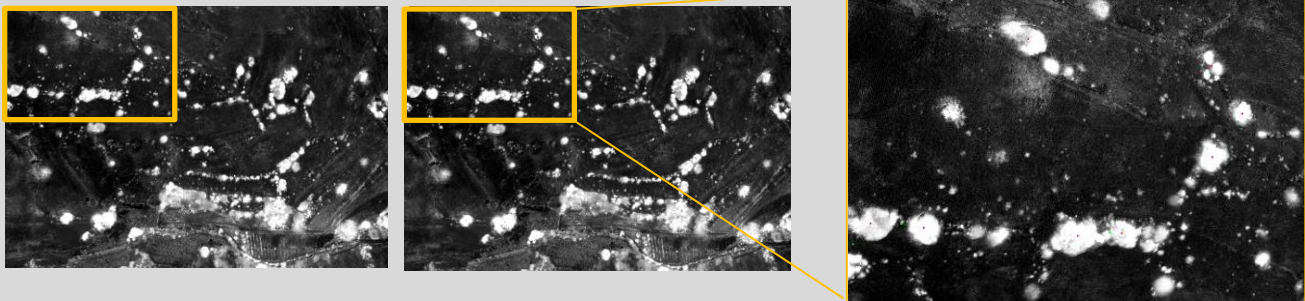


Fig. 1. Hexagonal grid method to classify image features.

Is worth to mentioned that and adaptation has been done to the grid apply the hexagonal method. The advantages of the method over others in the literature are basically three. First of all, NDVI image can discriminate on other information (not vegetation). Secondly, the resolution of multispectral images is superior to satellite images. Finally, actual method calculates not only maximum values, but also saddle points (points between two potential trees).

Results :

The method was tested in some aerial pictures taken at Oaxaca, Mexico. The images were of agricultural and some reforested areas ecological landscape. The image acquisition equipment consisted of an unmanned aerial vehicle (UAV): an eBee Sensefly model drone that has a Global Positioning System (GPS) integrated to it and a DJI Matrice carrying a Parrot Sequoia Multi-spectral/RGB camera. The Sequoia camera has four spectral cameras to capture the four spectral bands: Green (550 nm), Red (660 nm), Red Edge (735 nm), Near Infrared (790 nm), as well as the RGB camera.



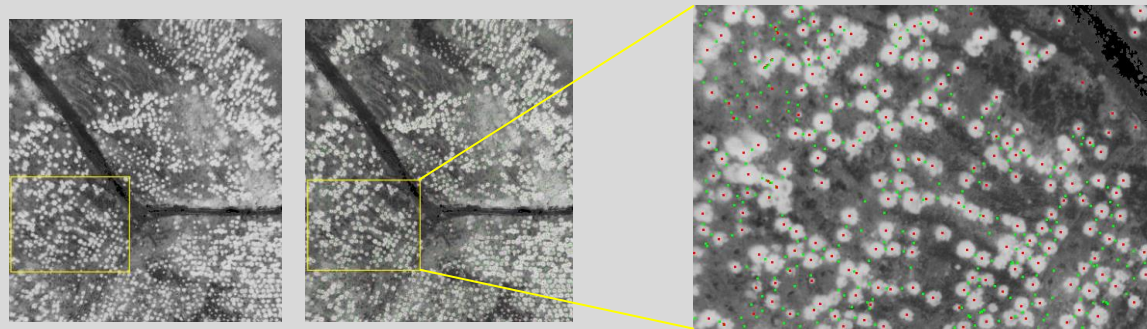
a) Original NDVI image

b) blobs and saddles detected

c) Segment showing trees detected (red dots)

Fig. 2. Agricultural land NDVI image with detected trees (blobs) at a fixed scale of $\sqrt{t} = \sigma = 25$ and a threshold of binarization of 0.4.

The figure 1 shows an agricultural land with sparse and separated trees, it was acquired at a height of 121 m of flight, at resolution of 12 cm per pixel, the size of the image is 1940 x 1326 pixels. The following image (fig. 2) is a reforested area near of dense forest in mountains, the size is 1280 x 960 pixels. It can be observed that detection at a certain scale the estimator can find some specific size of trees and other size trees can be found by varying scale parameter (sigma control parameter).



a) Original NDVI image b) blobs saddles detected c) Segment showing trees detected (red dots)

Fig. 3. Reforest land NDVI image with detected trees (blobs) at a fixed scale of $\sqrt{t} = \sigma = 5$ and a threshold of binarization of 0.4.

Conclusion and perspectives:

An estimator to detect trees in forest and ecological landscape was developed. Until now, not scale-space paradigm was mixed with UAV imagery to develop a tree density estimator applied to forest landscape using a VI image. Use of NDVI image was very useful to perform feature detection. Actual GIS and GEOBIA technologies are needed to work with innovative methods as described in this work. The method can be extended or improved by automatically calculating the space at full scale. Calculated saddle points can be considered as watersheds to separate two potential adjacent trees overlapping. Another possible optimization is the calculation of the diameter of the crown.

References :

- [1] Blaschke, Thomas. "Object based image analysis for remote sensing." *ISPRS journal of photogrammetry and remote sensing* 65.1(2010): 2-16.
- [2] Gomes, Ferreira and Philippe Maillard. "Detection of Tree Crowns in Very High Spatial Resolution Images." (2018).
- [3] Shin P, Sankey T, Moore MM, Thode AE. "Evaluating Unmanned Aerial Vehicle Images for Estimating Forest Canopy Fuels in a Ponderosa Pine Stand". *Remote Sensing*. 2018; 10(8): 1266.
- [4] Dralle Kim, Mats Rudemo. "Stem number estimation by kernel smoothing of aerial photos". *Canadian Journal of Forest Research* 26.7. pp. 1228-1236. (1996).
- [5] Guldogan, Olcay, et al. "Automated tree detection and density calculation using unmanned aerial vehicles." *Visual Communications and Image Processing (VCIP)*, 2016. IEEE, 2016.
- [6] Li, Wang, et al. "Remote estimation of canopy height and aboveground biomass of maize using high-resolution stereo images from a low-cost unmanned aerial vehicle system." *Ecological indicators* 67 (2016): 637-648.
- [7] Chianucci, Francesco, et al. "Estimation of canopy attributes in beech forests using true colour digital images from a small fixed-wing UAV". *International Journal of Applied Earth Observation and Geoinformation*. 47 (2016): 60-68.
- [8] T. Lindeberg, "Scale-Space". In Wiley Encyclopedia of Computer Science and Engineering. B. W. Wah (Ed.). 2008.
- [9] G. J. Hay, "A scale-space primer for exploring and quantifying complex landscapes". *Ecological modelling*. pp. 27-29. (2002).
- [10] J. Blom. Topological and Geometrical Aspects of Image Structure. PhD thesis, University of Utrecht, Department of Medical and Physiological Physics, Utrecht, The Netherlands, 1992.

