DOI: 10.19279/TVZ.PD.2017-5-1-02

PREPOZNAVANJE LICA – BRZIM PUTEM STROJNOG UČENJA DO POVEĆANJA UČINKOVITOSTI

FACE RECOGNITION – MACHINE LEARNING HIGHWAY TO EFFICIENCY

Milan Bajić[1], Matjaž Debevc[2]

[1]*Zagreb Univeristy of Applied Sciences,Zagreb, Croatia*
[2]*Faculty of Electrical Engineering and Computer Science Maribor, Maribor, Slovenia*

Sažetak

Prepoznavanje lica opisuje računalno upravljano prepoznavanje ljudskih lica u digitalnom mediju, slici ili videu. Iako započeta u šezdesetim godinama prošlog stoljeća, novi val istraživanja prepoznavanja lica danas bilježi sve veći porast interesa djelomično i zbog sigurnosne primjene te smanjivanje mogućih sigurnosnih ugroza.

U radu su dani smjerovi istraživanja u području, s naglaskom na korištene algoritme, skupove podataka za veću učinkovitost i pregled najčešće korištenih senzora. Jedna od novosti u području u posljednje vrijeme je korištenje različitih mehanizama strojnog učenja kao sredstva za povećanje učinkovitosti sustava za prepoznavanje lica. Predstavljanjem različitih autorskih pristupa dane su moguće smjernice za buduća istraživanja.

Ključne riječi: *prepoznavanje lica, strojno učenje, algoritam, sigurnost*

Abstract

Face recognition describes the automated [illegible] digital media, image or video. Even though started in the 1960's, the hype surrounding face recognition nowadays is more increasing, partially because of security applications it could provide and reduce possible security oriented threats. This paper is going to show directions of recent research in this field, with an accent on used algorithms, data sets for higher efficiency, and an overview of predominantly used sensors. One of the concept that is introduced lately in the field are various types of machine learning as means for creating more efficient face recognition systems. By presenting different approaches, possible directions are given for future research.

View metadata, citation and similar papers at core.ac.uk
brought to you by CORE

Keywords: *face recognition, machine learning, algorithm, security*

1. Uvod

1. *Intoroduction*

"As one of the most successful applications of image analysis and understanding, face recognition has recently received significant attention, especially during the past few years" [1]. This citation is from the year 2003, but summons all the arguments for doing another overview of recent research in the field. A literature overview in this paper will cover research results from 2015 till January 2017, within 3 areas: face recognition, machine learning within the face recognition domain and multisensory face recognition. The face recognition system is dominantly identified [illegible] automated machine image analysis for the purpose of face identification – matching query face with faces in a database, with the aim to associate the identity of the query with an identity in a database and face verification - which authenticates image query with claimed identity [2].
According to Brink, Richards, Fetherolf there are five advantages to machine learning over manual analysis, predefined rules and simple statistical models.

The ML (machine learning) can automatically increase accuracy of the decision-making engine with more data being collected. Through validation of answers, the ML models can learn new patterns automatically and the user can embed the ML in an automated workflow. By generating answers in milliseconds, the ML is fast and systems can react in real time. The ML is customizable, because the ML decision models are built from specific data and can be configured to optimize metrics required by a particular business. Scalability of the ML helps imagine a future with extreme amounts of data which can be handled with algorithms extracting knowledge from many machines in a cloud [3].

Face recognition, at first, is associated with a visible photo or a video sensors capturing light reflection from a surface. The limitations of using just a visible spectrum are many, some of which are, shadows over areas of interest, over or under exposed areas, variable lighting conditions in different time of a day, change in facial expression, and nighttime surveillance [4]. Chen, Flynn, and Bowyer [5] found that the combination of IR (infrared) and a visible light image can outperform either of them alone by means of face recognition. By applying band selection algorithms to find most discriminative bands in a visible and the NIR spectrum, and information fusion from different sensors, Uzair, Mahmood and Mian [6] found this a promising approach.

2. Face recognition

2. *Prepoznavanje lica*

According to Moses et al. [7] "the variations between the images of the same face due to illumination and viewing direction are almost always larger than image variations due to change in face identity." One approach that solves this problem is proposed by introducing a face normalization algorithm in the face recognition process in an unconstrained environment [8]. Still, the authors give a limitation of their algorithm, it works much better in normalizing near frontal face pose, while distorting those semi profile and negatively impacting recognition accuracy. Philips et al. [9] add to this growing variability algorithm ethnics origin, as well as data set ethnics mix with a conclusion that "the performance of state-of-the-art face recognition algorithms varies as a joint function of the demographic origin of the algorithm and the demographic structure of the test population". To solve this multiple variability issue in face recognition process in real life situations, one should try new approaches and develop new pre-processing and post-processing algorithms, or create completely new processes of face recognition to bridge the biggest gaps in data fusion from various sources and achieve better recognition results from known and well researched sensors. Mutual Component Analysis [10] is introduced as a tool for heterogeneous face recognition or matching two face images from alternative image modalities. The author's stages of development are "generative model is first proposed to model the process of generating face images in different modalities, and then an Expectation Maximization (EM) algorithm is designed to iteratively learn the model parameters." This model can infer mutual components unique to a person's identity and enables fast and efficient matching for cross modality face recognition. MCA applied in experiment [10] gives results that can significantly outperform state of the art in infrared – visible modality FR.

In their paper, "Heterogeneous Face Detection," [11] the author's work on accuracy improvement of existing FR algorithms by applying various image enhancement techniques before applying the Viola Jones FR algorithm on thermal and infrared images of the crowd. Infrared images are mostly low resolution, and in their experiment, they apply histogram equalization – "image enhancement technique through which the grey levels are spread out so they are evenly distributed across their range." Thermal images are poor quality, distorted and unclear, but the biggest problems are lack of texture and not well defined edges. So, the proposed solution [11] is use of YCbCr (luma component, blue difference, red difference channels) colour model and contrast limited adaptive histogram equalization (CLAHE) to make detection of faces easier. Poon, Amin, Yan [12] used the GRF – Gradientfaces algorithm for normalization of images, illumination in pre-processing stage of FR to enhance recognition rate.

The GRF algorithm works only on grayscale images, so all the colour images had to be converted into that mode. This algorithm was picked up for their [12] PCA based FR system as one working best; and results of recognition are improved on their datasets from 6.25% up to 60.75% for varying illumination, and from 38% to 99.50% for colour images with colour background.

To have application in surveillance, access control or identity authentication FR system should be adjusted for work in real life environment, where acquired images are not made in a perfect environment. The authors of "FACE RECOGNITION IN REAL-WORLD IMAGES" [13] present results that outperform state-of-the-art methods on real world face identification. By applying a set of algorithms, they first automatically align a face to prepare it for identification. Identification is made by the RSC – Robust Sparse Coding algorithm which is an improved version of sparse representation method, less prone to occlusions and lighting changes. A paper on face recognition based on person specific identification [14] is done on FR and person identification from surveillance video camera. Video with its own disadvantages in variability provides more information than still images do [14]. Working on video, on the other side, introduces new problems of extremely high amounts of data for processing and putting high demands on computational resources, even more on real time FR and identification. The author's approach is generating still images from a video, applying the Viola Jones algorithm for face detection, clustering by K-means and FR using the HOG (Histogram of Oriented Gradients) and the LPB (Local Binary Pattern) feature extraction, and combining those features resolves occlusions problems much better.

3. Machine Learning
3. *Strojno učenje*

While applying convolutional neural networks (CNN) on certain image data sets has become almost 100% accurate [15], new more challenging data sets are developed, thus it will take much more time to mark this issue as solved.

State of the art and open source FR systems found much more challenging working on new, unconstrained data sets, "The term "unconstrained" implies a system can perform successful identifications regardless of face image capture presentation (illumination, sensor, compression) or subject conditions (facial pose, expression, occlusion)" [16]. With their big numbers FR challenge "The Mega Face Benchmark: 1 Million Faces for Recognition at Scale" [17] results, in using classical algorithms working on large data sets performance degrades, difference in performance of FR solutions is much more visible than on small scale data sets.

In machine learning, FR systems can be negatively influenced by incorrect annotations and low quality images [18], but important information for system accuracy are hidden in the correct landmark points and regions [15], which are mostly unavailable for public images. Ye et al. [18] propose active annotation and learning framework for work with FR models on large-scale dataset with noisy data and insufficient annotations. In their paper "A Convolutional Neural Network Cascade for Face Detection" [19] Li et al., propose CNN cascade for fast detection of faces in early stages of pre-processing at lower resolution thus reducing the areas of interest for full resolution computing in later stages.

Even tough FR from still visual images are well researched and described in literature, video FR or VFR in CNN needs specific training [20] to target blurriness, occlusion and dramatic pose variation. Many authors [20] argue for artificial blurriness on training sets to match real life video situations, for occlusion and pose variation, they suggest their model of TBE-CNN (trunk-branch ensemble CNN) where trunk network is trained to learn face representation for holistic face images, while each branch is trained to learn face representation for image patches cropped from one facial component. CNN can also be used in Near Infrared face recognition (NIR) with similarly good results [21], although the NIR data sets are small and not particularly efficient for learning systems which are most accurate as learning data sets become bigger so development of these sets is as important as development or improvement of algorithms.

4. Multi sensor face recognition

4. *Multisenzorsko prepoznavanje lica*

Visual sensors are mostly used in FR systems together with their limitations, as sensors prices go lower and new sensors are becoming commercially available they also become part of the research. Low quality, low cost Microsoft Kinect Sensor [22], is one sensor that brings opportunity for research of 3D FR over very expensive and unreachable 3D systems. This kind of devices are not intended to be used as FR sensors, so data had to be registered in order for image RGB data to be matched with the depth data. Kinect is a multi sensor device [23] consisting of IR laser emitter, IR camera and RGB camera, but the data acquired are only matched for intended purpose. Second, the multi sensor device is Intel RealSenseTM, this sensor has 3 RGB cameras and stereoscopic IR for depth maps [23]. Not all the research is done on integrated shelf multi sensor devices. WU et al. [24] advocate usage of visible and thermal images for facial expression recognition. Thermal images, in their case, are just used in a training phase of automatic classification, where two networks are trained separately, but after the training phase only visible images were used in classification. "The selection of appropriate sensors has become one of the most significant factors for high-performance vision systems" [25], and creating useful integrated data from them could be one of the biggest challenges.

The authors of the paper "3D Multi-Spectrum Sensor System with Face Recognition" created a system with two subsystems: (1) a 3D multi spectrum sensor system, and (2) a face recognition system. The 3D sensor takes 0.07 sec. to generate 3D multi spectral data and the face recognition system processing time is less then 1 msec., enabling it for use in real time FR applications, such as a robot and surveillance.

5. Diskusija

5. *Discussion*

Visible, infrared, thermal IR, multispectral, hyperspectral, 3D, time of flight camera, pre-processing, processing post processing, algorithm innovation, algorithm improvement, and machine learning, each of these areas or sensors can influence FR accuracy. While machine learning is recently one of the most accented research in that domain, it can work only on big data sets which do not exist for some of the sensors. The multi sensor approach brings challenge of proper data fusion from various sensors as well as extraction of most useful data from each sensor. Face recognition in unconstrained environments is more influenced by pose or illumination of face, then change in identity – which can cause serious amount of false alarms as real time application are almost always done in unconstrained environment.

Ethnicity of algorithm is one of the FR problems because training sets used can influence accuracy on other data sets, thus efficiency declared by different authors needs to be double checked and verified in situations where algorithms are to be applied. While pose and illumination changes are the biggest challenges in FR, there are solutions that can help optimize process accuracy by interventions in each stage from data acquisition, over preprocessing all the way to face identification or recognition. Most authors propose a use of well tested and recognized algorithms with a modification to prepare images to make the work easier thus reducing computing time or needed resources.

Machine learning makes traditional approach to FR almost 100% accurate, on certain regularly used data sets. On the other hand, machine learning brings new challenges like working on unconstrained images – those that are acquired in regular, non-perfect conditions. To enable higher quality research on machine learning FR, various high quantity publicly available visible image data sets for training of system are available. Not so with other sensors, infrared, probably second most used sensor does not have such large data sets, so training is not done properly and it can cause more false alarms and lower accuracy. Machine learning systems for FR can be strengthen by adding additional information about images, place, location, date thus raising accuracy. Faster systems are trained by reduction of areas of interest on high resolution images and significantly reduces time of processing.

To use effectively the ML in FR system with multiple sensors there must be training data sets with enough information for each sensor. The multi sensor FR system can work with of the shell multisensory devices like Kinect or Intel RealSenseTM, which makes data integration and fusion from all used sensors easier. Another option is to create hand crafted systems from individual sensors, with the purpose for use in mind, but this requires additional work on proper data integration and fusion. These systems are most promising crafted solution for specific purposes, but it will take more time to test them and make them ready for work. Real time FR can be achieved by using a multiple sensor based machine learning system trained on large scale data sets, with computational speed under 1 sec.

The next step would be a creation of large unconstrained image sets for each used sensor. One possible solution for illumination diversity problem could be the use of the HDR (High Dynamic Range) visible camera which makes 5 different exposures of the same scene and automatically merges them in one evenly exposed image. The applications are countless and new ones will arise when these kinds of systems become normal in everyday use for surveillance, homeland security or border control.

6. Reference

6. *Refrences*

[1.] Zhao, W., et al., Face recognition: A literature survey. ACM computing surveys (CSUR), 2003. 35(4): p. 399-458.

[2.] Jain, A.K. and S.Z. Li, Handbook of face recognition. 2011: Springer.

[3.] Brink, H., J. Richards, and M. Fetherolf, Real-world machine learning. 2014: Manning.

[4.] Bourlai, T., Face recognition across the imaging spectrum. 2016: Springer.

[5.] Chen, X., P.J. Flynn, and K.W. Bowyer, IR and visible light face recognition. Computer Vision and Image Understanding, 2005. 99(3): p. 332-358.

[6.] Uzair, M., A. Mahmood, and A. Mian, Hyperspectral face recognition with spatiospectral information fusion and PLS regression. IEEE Transactions on Image Processing, 2015. 24(3): p. 1127-1137.

[7.] Moses, Y., Y. Adini, and S. Ullman, Face recognition: The problem of compensating for changes in illumination direction. Computer Vision—ECCV'94, 1994: p. 286-296.

[8.] Haghighat, M., M. Abdel-Mottaleb, and W. Alhalabi, Fully automatic face normalization and single sample face recognition in unconstrained environments. Expert Systems with Applications, 2016. 47: p. 23-34.

[9.] Phillips, P.J., et al., An other-race effect for face recognition algorithms. ACM Transactions on Applied Perception (TAP), 2011. 8(2): p. 14.

[10.] Li, Z., et al., Mutual component analysis for heterogeneous face recognition. ACM Transactions on Intelligent Systems and Technology (TIST), 2016. 7(3): p. 28.

[11.] Das, A., R.K. Kumar, and D.R. Kisku. Heterogeneous Face Detection. in Proceedings of the International Conference on Internet of things and Cloud Computing. 2016. ACM.

[12.] Poon, B., M.A. Amin, and H. Yan, PCA Based Human Face Recognition with Improved Methods for Distorted Images due to Illumination and Color Background. IAENG International Journal of Computer Science, 2016. 43(3).

[13.] Fontaine, X., R. Achanta, and S. Süsstrunk. Face Recognition in Real-world Images. in IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2017.

[14.] Kokila, S. and B. Yogameena. Face Recognition Based Person Specific Identification for Video Surveillance Applications. in Proceedings of the Third International Symposium on Women in Computing and Informatics. 2015. ACM.

[15.] Ding, C. and D. Tao, Robust face recognition via multimodal deep face representation. IEEE Transactions on Multimedia, 2015. 17(11): p. 2049-2058.

[16.] Klare, B.F., et al. Pushing the frontiers of unconstrained face detection and recognition: IARPA Janus Benchmark A. in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015.

[17.] Kemelmacher-Shlizerman, I., et al. The megaface benchmark: 1 million faces for recognition at scale. in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016.

[18.] Ye, H., et al. Face Recognition via Active Annotation and Learning. in Proceedings of the 2016 ACM on Multimedia Conference. 2016. ACM.

[19.] Li, H., et al. A convolutional neural network cascade for face detection. in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015.

[20.] Ding, C. and D. Tao, Trunk-branch ensemble convolutional neural networks for video-based face recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017.

[21.] Peng, M., et al., NIRFaceNet: A Convolutional Neural Network for Near-Infrared Face Identification. Information, 2016. 7(4): p. 61.

[22.] Ajmera, R., A. Nigam, and P. Gupta. 3d face recognition using kinect. in Proceedings of the 2014 Indian Conference on Computer Vision Graphics and Image Processing. 2014. ACM.

[23.] Freitas, T.D.S., 3D Face Recognition Under Unconstrained settings using Low-Cost Sensors. 2016.

[24.] Wu, C., et al. Facial Expression Recognition with Deep two-view Support Vector Machine. in Proceedings of the 2016 ACM on Multimedia Conference. 2016. ACM.

[25.] Kim, J., et al., 3D Multi-Spectrum Sensor System with Face Recognition. Sensors, 2013. 13(10): p. 12804-12829.

AUTORI · *AUTHORS*

Milan Bajić - nepromjenjena biografija nalazi se u časopisu Polytechnic & Design Vol. 4 No. 4, 2016.

Korespodencija
milan.bajic@tvz.hr

Matjaž Debevc

Dr. Matjaž Debevc received the B.S., M.S. and Ph.D. degrees in computer science from University of Maribor, Faculty of Electrical Engineering and Computer Science. He is currently Associate Professor in the field Computer Science and Control Engineering. His research interests include e Learning, human-computer interaction, user interface design, adaptive user interfaces, internet applications, virtual reality and applications for disabled people. He participated as leader and partner in more than 50 EU and national research and development projects. For his work in e-Learning, he was, in 2007 and 2011, the recipient of the Comenius EduMedia Medaille Award from GPI-Gesellschaft für Pädagogik und Information e.V. for didactical multimedia product "ECDL E-learning materials" and “How to get a job” for deaf and hard of hearing people.He has published more than 350 articles, papers, books and publications for scientific journals and conferences. Dr. Debevc is a general chairman of IEEE Education Society Chapter in Slovenia and a member of ACM. Since 2016 he is a member of the Expert Group on the Accessibility at World Federation of the Deaf (WFD).

Korespodencija
matjaz.debevc@um.si