A THEORETICAL MODEL FOR COMPUTER-AIDED IMAGE RECOGNITION SYSTEM BASED ON DIVERSE INPUTS

Richa Sharma[1] and T. R. Gopalakrishnan Nair[2]
[1]Research Scholar, Jain University, Bangalore, India
[2]Rector, Raja Rajeshwari College, Bangalore, India

ABSTRACT

This paper proposes a multiple senses model to match with the object/pattern recognition accuracy, which the human being achieve, with the help of five sense organs, against the machines. The proposed model can help the computer-aided pattern recognition to reach as close as possible to natural visual perception. The initial investigation based on 16 X-ray images shows that even if the input from a single source (X-ray machine) is exploited in the best possible way, using spectral analysis (Fourier, wavelets, morphological etc.), geometry, fractals, statistics etc. The classification accuracy of an object/pattern can further be improved considerably if we effectively fuse inputs received from other diverse sources for the same object. An effective fusion of information, collected from diverse sources, can make computer-aided image recognition closer to natural visual perception. The initial investigation based on 16 X-ray images affected with Tuberculosis shows that the disease recognition accuracy got improved from 85% to 93 % when we added two additional features from a different source of input.

KEYWORDS: *Image Recognition, Pattern Detection, Computer aided Detection, Similarity Measures*

I. INTRODUCTION

In many aspects of our life, we have successfully replaced human with machines and the machines are doing a much better job. Machines have effectively substituted human when it comes to strength or power related jobs like building/road construction, transportation, etc. We also rely on machines for faster numerical calculation, storage, and fast retrieval. However, when it comes to learning and intelligence, machines are still lacking human capabilities in terms of speed of learning, prediction, and interpretation. For example, image processing and machine learning applications like face/object/gesture/age recognition fail measurably at times when the illumination is changed, or a part of the object is hiding behind the other objects.

When intelligent algorithms are applied on digital images using a digital computer to detect the object of interest in an automatic way, it can be called Computer Aided Detection (CAD). It is expected from a face recognition system that it must recognise a face from an original image despite all the variations between images of the same face. A common approach to overcome image variations due to changes in the illumination is to use image representations that are relatively insensitive to these variations. Examples of such representations are edge maps, image intensity derivatives, and images convolved with 2D Gabor-like filters. It is found that none of the representations considered is sufficient by itself to overcome image variations. Similar results were obtained for changes due to variation in viewpoint and expression. Humans perform considerably better under the same conditions [2,13]. Heavy research is going in the direction of computer aided detection and recognition to handle such limitations [1-3,15,16].

If we look at the state of the art Video Management Systems (VMS), we find that even the latest versions of the leading video analytics companies like Milestone, Genetic, AllGoVision, etc. do not perform as good as humans. For applications like smoke or fire detection, people counting, face recognition, missing object or left object detection, despite using latest techniques and algorithms, the accuracy of such applications are not on par with humans. Even though all the leading VMS companies claim accuracy for such applications between 70-95%, they are tested under limited conditions [12]. For example, in case of face recognition in video surveillance systems, faces are tested only when the person is facing the camera under proper illumination, and the face is covering more than a specific size of the camera view. Faces go unrecognised when a person stays far from the camera or if he moves fast, i.e. insufficient duration of appearance in the camera view. When it comes to humans, we can easily recognise faces at varying distances with different view and expressions. Even if half the face is covered, or hairstyle is changed, we can easily recognise the person correctly whereas the CAD/ VMS systems may fail miserably. Heavy research is in vogue in the direction of computer aided detection and objects recognition to handle such limitations [1-3,14].

This paper proposes a multiple senses model to match with the object/pattern recognition accuracy, which the human being achieve, with the help of five sense organs, against the machines. The theoretical model for the technique used in the current approach is presented in section 2. The background of similarity measures is explained in section 3. Section 4 discusses the experimental results which demonstrate the effective use of similarity measures and diverse inputs using the proposed method. Finally, we present the concluding remarks which involve the highlight of this work, and future work.

II. THE PROPOSED APPROACH

The reason for machines not as efficient as the human in doing intelligent tasks is, we, the humans are enabled with five sense organs to receive the inputs. The brain combines all the five inputs and concludes, but the traditional Image Recognition Systems and VMS rely on inputs received only from the cameras (videos or pictures). It is not possible to match with the accuracy which the human being achieve with the help of five sense organs, against the machines those are using just one input device. Hence, if we want machines to do a better job, there is a strong need to generate and provide more inputs to the computer. We propose a multiple senses model which can certainly make the computer-aided pattern recognition as close as possible to natural visual perception.

2.1 Multiple Senses Theoretical Model

When a child is born, he starts receiving inputs from his/her five sense organs and the training starts. He receives (verbal, visual and sensory) inputs from the environment and starts keeping a combination of these three types of inputs in his memory along with the recognition/classification results. By the time he grows big, he has a well-developed and well-trained brain which can help him to detect/ recognise/ predict even unknown things.

In order to make the computer-aided pattern recognition closer to natural visual perception, we need to do the following -

Step 1. Have inputs from various possible dimensions depending on the application (Videos/ Pictures, Sound (microphone), Temperature etc.) and generate a collective feature set based on these diverse inputs. Some meta data providing input device's installation details, like the height of the camera, camera view, scene details (indoor/ outdoor), temperature, noise, resolution etc. can also be given as a part of the input depending on the problem. The metadata and inputs depend entirely on the type of application. For example, if it is fire detection VMS, the temperature of the surrounding should be one of the major inputs along with the camera feed. We used a similarity measure (as explained in Section 3) in the process of generating a collective feature set.

Step 2. Have a CAD system setup with ANNs, genetic algorithms or other equivalent techniques for the learning part. Initial training may require human intervention.

In case of computer aided diagnostics of medical images, the same approach can be applied. To get a better understanding of the critical cases, we can get inputs from all possible sources like ultrasound,

X-rays, MRI, pathology and clinical symptoms. Once inputs form these diverse sources are collected, we can generate a similarity measure based on the following approach.

III. SIMILARITY AND SIMILARITY MEASURES

In our brains, we tend to store the data in terms of similarity and differences. The similarity plays a substantial role in discrimination of textures. In general, it is a formal bi-variable reciprocal and symmetrical relation described in a set of objects [4]. It may not satisfy a transitivity condition always. That is, if an object 1 is similar to object 2 and object 2 is similar to object 3, then not obviously object 1 is similar to object 3; this, in particular, can be proven if 1, 2 and 3 denote samples of textures. On the other hand, if we partition same texture into sub-areas, then any three samples taken from them satisfy not only the reciprocity and symmetry but also the transitivity of similarity conditions. This apparent contradiction can be overcome by using a concept of similarity measure [4]. This section discusses the definitions used for generating the Similarity Measures; the detailed proof of the similarity approach can be seen in [4].

Definition 1. Let Ω denote any set consisting of more than two elements (objects). We call similarity measure a function σ described on a Cartesian product Ω^2 satisfying the conditions:

i. $0 <= \sigma(\omega', \omega'') <= 1$,

ii. $\sigma(\omega', \omega') = 1$,

iii. $\sigma(\omega', \omega'')=\sigma(\omega'', \omega')$,

iv. $\sigma(\omega', \omega'').\sigma(\omega'', \omega''') <= \sigma(\omega', \omega''')$ for any ω', ω", ω''' belongs to $\Omega\bullet$

The condition iv reminds a well-known "triangle inequality" in a definition of distance measure in a metric space [6]. If Ω is also a metric space and d(ω', ω") denotes a distance measure between any two its elements, then their similarity measure can be defined as

$$\sigma(\omega', \omega'')=\exp[-\alpha.d(\omega', \omega'')] \quad (1)$$

Where α is a positive scaling coefficient. It can easily be proven that the conditions i-iv of Definition 1 are then satisfied; in particular, iv is satisfied due to the inequality:

$$d(\omega', \omega''') <= d(\omega', \omega'') + d(\omega'', \omega''') \quad (2)$$

The set Ω with a defined in it similarity measure σ will be called a similarity space. On

the basis of Definition 1 it can be formulated:

Definition 2. Let Ω be a similarity space. Any non-empty subset Sε □Ω such that $0 <\varepsilon<1$ and any two of its elements ω', ω" ϵ Sε satisfy the condition $\sigma(\omega', \omega'') >=\varepsilon$ will be called an ε -similarity

class in $\Omega\bullet$.

Evidently, in any ε-similarity class Sε, the ε-similarity of its elements is transitive by definition.

IV. SIMILARITY MEASURES FOR DISCRIMINATION OF IMAGE TEXTURE

A single pixel in any medical image is not intended to represent a single cell of the organ. Nevertheless, when a collection of cells changes, due to infection or any disease, the change can be visualised in the form of a change in the hue of the pixels, in the digital medical images. For Example, in the spine X-ray images, destruction of the intervertebral disk space and the adjacent vertebral bodies comes out in the form of unclear- hazy boundaries. Healthy vertebra shows clear boundaries in X-ray images. We used this observation to derive relevant features to describe the normal and abnormal cases.

We studied 16 X-ray images of the Lumbar Spine. All images are enhanced by mapping the gray values such that 1% of data are saturated at low and high intensities. To sharpen the features and flatten the lighting variations in the image, a homomorphic wavelet filter with a double-level decomposition is applied. The Region Of Interest (ROI) of size 128*128 pixels are cropped covering TB affected inhomogeneous regions (unclear- hazy boundaries of the spine) and saved in a dataset named 'TB'.

Normal regions of the healthy Lumbar spine X-ray images were also cropped to the same size and stored in a dataset named 'NOTB'. All the images are cropped under the supervision of a qualified doctor.

Feature Vectors (FV) are extracted from the TB and NOTB images. Feature vectors were formed with eight texture feature descriptors of the ROI [7-8], and ten Gray Level Difference Matrix (GLDM) features as explained in [9] and three descriptors are taken from Gabor plot as they were found to be relevant to our previous work of a similar nature [10,11]. The eight texture feature descriptors are- Average gray level, Average contrast, Measure of smoothness, Third moment, Measure of uniformity, Entropy, Energy, and Maximum column sum energy. Table 1 shows the details of all the feature descriptors used. We used FV3 for testing of the proposed method.

Table 1. List of Feature Descriptors used

Name of the Feature Set	Feature Descriptors	Size
FV1	Texture Features: Average gray level, Average contrast, Measure of smoothness, Third moment, Measure of uniformity, Entropy, Energy, Maximum column sum energy, and data 10 to 19 of GLDM of the ROI	8 + 10 = 18 values
FV2	FV1, data 1 and data 2 of GLDM of Gabor Phase plot	18 + 2 = 20 values
FV3	FV2, Sum of entire data of Gabor Amplitude plot	20 + 1 = 21 values

All the image feature descriptors were normalised and brought to the same scale before feature set generation using mean normalisation. Mean normalisation is done to standardise the range of independent variables or features of data.

A similarity matrix was generated (using equation (1) with $\alpha = 0.01$) based on the feature vectors (FV3), generated from the TB affected and normal images. Euclidian distance is used to calculate the distance between two feature vectors d(ω', ω"). Table 2 presents the similarity matrix using seven TB and six normal (NOTB) images.

Table 2. Similarity Matrix

Image No	TB 1	TB 2	TB 3	TB 4	TB 5	TB 6	TB 7	NOTB1	NOTB2	NOTB3	NOTB4	NOTB5	NOTB6
TB 1	1	0.4800	0.7178	0.2476	0.5287	0.4158	0.2959	0.0003	0.0606	0.1029	0.0007	0.2841	0.0003
TB 2	0.4800	1	0.4802	0.2616	0.7922	0.3975	0.2952	0.0004	0.0722	0.1142	0.0008	0.3280	0.0003
TB 3	0.7178	0.4802	1	0.3396	0.4853	0.5024	0.4074	0.0005	0.0827	0.1404	0.0009	0.3847	0.0004
TB 4	0.2476	0.2616	0.3396	1	0.2325	0.3712	0.7589	0.0013	0.2424	0.4095	0.0026	0.6580	0.0010
TB 5	0.5287	0.7922	0.4853	0.2325	1	0.3535	0.2614	0.0004	0.0623	0.1007	0.0007	0.2795	0.0003
TB 6	0.4158	0.3975	0.5024	0.3712	0.3535	1	0.4759	0.0006	0.1072	0.1659	0.0012	0.5036	0.0005
TB 7	0.2959	0.2952	0.4074	0.7589	0.2614	0.4759	1	0.0011	0.1988	0.3277	0.0021	0.7578	0.0008
NOTB1	0.0003	0.0004	0.0005	0.0013	0.0004	0.0006	0.0011	1	0.0054	0.0032	0.4024	0.0011	0.6380
NOTB2	0.0606	0.0722	0.0827	0.2424	0.0623	0.1072	0.1988	0.0054	1	0.5563	0.0107	0.1959	0.0040
NOTB3	0.1029	0.1142	0.1404	0.4095	0.1007	0.1659	0.3277	0.0032	0.5563	1	0.0064	0.3044	0.0024
NOTB4	0.0007	0.0008	0.0009	0.0026	0.0007	0.0012	0.0021	0.4024	0.0107	0.0064	1	0.0021	0.2636
NOTB5	0.2841	0.3280	0.3847	0.6580	0.2795	0.5036	0.7578	0.0011	0.1959	0.3044	0.0021	1	0.0008
NOTB6	0.0003	0.0003	0.0004	0.0010	0.0003	0.0005	0.0008	0.6380	0.0040	0.0024	0.2636	0.0008	1

Table 3. Average Range of Similarity Measure for each category

Mean (TB vs TB)	Mean (NOTB vs NOTB)	Mean (TB vs NOTB)	Mean (NOTB vs TB)
0.45-0.53	0.25-0.34	0.07-0.21	0.0005-0.46

It is evident in Table 2 that the similarity measure is high usually between any two TB images and goes low between any two TB and NOTB images. However, it is not the case always. Hence, as per our model, we need to give more inputs from diverse sources like ultrasound, X-rays, MRI, pathology and clinical symptoms, personal details like age, sex, weight etc. In this case, we had only clinical symptoms available with us. We gave two clinical symptoms; history of fever and weight loss as additional feature inputs.

V. RESULT AND DISCUSSION

There are several ways the similarity matrix and additional feature inputs can be utilised; we used weak similarity function [4] which is defined as a weighted sum:

$$G(T;V)=\Sigma\ T_i * V_i \quad (3)$$

where $V= [V_1,V_2,...,V_n]$ is a vector of non-negative weights whose sum equals 1, assigning relative importance levels to the corresponding logical test's values $T_1, T_2,..., T_n$.

In stating similarity between two textures, two situations should be considered. The first one arises when the properties of one texture (a reference texture) are a priori given, and the similarity to it of the other one is to be stated. The second one arises when the properties of both textures are a priori not given, and the problem consists in stating their similarity or dissimilarity. In medical applications, both cases may arise. We used it for 2 class classification purpose, the properties of both classes (TB and NOTB) are a priori given, and the problem consists in assigning the new external sample to one of the two classes based on similarity or dissimilarity.

As a basis for description of a TB affected tissue ω', normal tissue ω" and a new external tissue ω"', we took the following quality features based on the values obtained in Table 3-

T_1: Mean Similarity measure using 21 point FV with TB σ(ω', ω"') - 0.5+-5%

T_2: Mean Similarity measure using 21 point FV with NOTB σ(ω", ω"') - 0.1+-10%

T_3: History of Fever

T_4: History of Weight Loss

For healthy/ill textures discrimination, a composed measure will be defined as:

$$Result=0.3T_1 + 0.3T_2 + 0.2T_3 + 0.2T_4 \quad (4)$$

A decision assigning an examined sample ω"' of texture to the class TB (similar to ω') will take the form:

ω"' ϵ TB if Result >= γ

ω"'does not belong to TB otherwise, (5)

where $0 < \gamma <= 1$ is a fixed threshold level. We used $\gamma=0.5$ for testing the images in databases TB and NOTB. The accuracy of classification using Test values T_1 and T_2 extracted from a single input (X-ray images) got improved from 85% to 93 % when we added two additional Test values T_3 and T_4 from a different source of inputs.

VI. CONCLUSION AND FUTURE WORK

The initial investigation based on 16 X-ray images shows that computer-aided image recognition accuracy using combinations of various qualitative features and numerical parameters extracted from a

single source can be considerably improved by taking inputs from diverse sources. The proposed model has improved the accuracy of classification, extracted from a single input (X-ray images), from 85% to 93 % by adding two additional features from a different source of inputs.

In the future, the classification accuracy can further be improved considerably if we effectively fuse inputs received from other diverse sources (MRI, pathology etc.). An effective fusion of information, collected from diverse sources, can make computer-aided image recognition closer to natural visual perception.

REFERENCES

[1]. Ralph Gross & Vladimir Brajovic (2003) An Image Preprocessing Algorithm for Illumination Invariant Face Recognition, Audio- and Video-Based Biometric Person Authentication, Vol. 2688 of the series Lecture Notes in Computer Science, pp 10-18., Springer Berlin Heidelberg Publisher, DOI: 10.1007/3-540-44887-X_2.

[2]. Y, Adini. Rehovot, Israel. Y, Moses & S, Ullman (1997) "Face recognition: the problem of compensating for changes in illumination direction", IEEE Transactions on Pattern Analysis and Machine Intelligence,Vol.19, Issue 7, pp 721 – 732, 10.1109/34.598229.

[3]. M. De Marsico, M. Nappi, D. Riccio & H. Wechsler (2013) "Robust Face Recognition for Uncontrolled Pose and Illumination Changes", IEEE Transactions on Systems, Man, and Cybernetics: Systems, Vol. 43, no. 1, pp 149-163, doi: 10.1109/TSMCA.2012.2192427.

[4]. Juliusz L. Kulikowski, Małgorzata Przytulska & Diana Wierzbicka (2011) "Discrimination of Biomedical textures Based on Logical Similarity Measure", Journal of Medical Informatics and Technologies, Vol. 17/2011, ISSN 1642-6037.

[5]. Kulikowski J.L.(2002) "From pattern recognition to image interpretation", Biocybernetics and Biomedical Engineering, vol. 22, no 2-3, pp 177-197.

[6]. Przytulska, M (2010) Report N N518 4211 33 of the Project on "Methods of computer analysis of radiological images for pathomorphological lesions assessment in selected inner body organs" (in Polish). IBBE PAS, Warsaw.

[7]. Noor, Norliza Mohd. Rijal, Omar Mohd. Ashari, Yunus. Mahayiddin, Aziah A. Peng, Gan Chew. Abu-Bakar, SAR. (2010) "A statistical interpretation of the chest radiograph for the detection of pulmonary tuberculosis" IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES). Kuala Lumpur, pp 47-51. DOI: 10.1109/IECBES.2010.5742197.

[8]. Patil, SA. Udupi, VR. Kane, CD. Wasif, AI. Desai, JV. Jadhav, AN (2009) "Geometrical and texture features estimation of lung cancer and TB images using chest X-ray database", International Conference on Biomedical and Pharmaceutical Engineering, Singapore, pp 1-7. DOI: 10.1109/ICBPE.2009.5384113.

[9]. Kim, Jong Kook. Park, Wook Hyun (1999) "Statistical texture features for detection of micro calcifications in digitized mammograms" IEEE Transactions on medical imaging, Vol 18, No 3, pp 231-238. DOI: 10.1109/42.764896.

[10]. Sharma, Richa. Nair, TR Gopalakrishnan (2016) "Studies of suspicious lesions through local texture analysis in spine radiographs" International Journal of Computer Application. Foundation of Computer Science. 146(2), pp 35-40. ISBN : 973-93-80893-78-2.

[11]. Nair, TR Gopalakrishnan. Sharma, Richa (2014) "Pre-transplant visualization of combined images for predictive medical analyses" Journal of Medical Engineering and Technology, Taylor & Francis publisher, 38(4), pp 220-226.

[12]. Website: https://www.milestonesys.com/contentassets/0537a5be61ed45328d3d6c473a1f58d09/ hp-autonomy-face-recognition-user-guide.pdf [accessed on 30/8/2017]

[13]. Website: https://www.parc.com/services/focus-area/video_image_analytics/[accessed on 30/8/2017]

[14]. Ng, Choon-Ching. Yap, Moi Hoon. Cheng, Yi-Tseng. Hsu, Gee-Sern (2017) "Hybrid Ageing Patterns for face age estimation" Journal of Image and Vision Computing, in Press.

[15]. Kayaa, Heysem. Gürpınar, Furkan. Ali Salah, Albert (2017) "Video-based emotion recognition in the wild using deep transfer learning and score fusion" Journal of Image and Vision Computing, Vol. 65, pp 66-75. Applied Soft Computing

[16]. Chaudhry, Shonal. Chandra, Rohitash (2017) "Face detection and recognition in an unconstrained environment for mobile visual assistive system" Journal of Applied Soft Computing, Vol. 53, April 2017, pp 168-180.

AUTHORS

Richa Sharma has 12 years of teaching, research and Industry experience. Recently, she worked for a Video Analytics Company named AllGoVision. She holds the M.Tech. degree (I.I.T. BHU, Varanasi) and is engaged in her doctoral program at Jain University, Bangalore. Her areas of interests are digital image processing and medical image analysis. She is a Member of International Association of Computer Science and Information Technology and Computer Society of India.

Dr. T.R. Gopalakrishnan Nair has 30 years of experience in the professional field spread over the industry, research and education. He holds degrees M.Tech. (I.I.Sc., Bangalore) and Ph.D. in Computer Science. He is currently the Saudi Aramco Endowed Chair of Technology and Information Management, in Prince Mohammad Bin Fahd University. He is the winner of PARAM award for technology innovations.