Colorimetric Recognition for Urinalysis Dipsticks Based on Quadratic Discriminant Analysis*

Yong He, Kai Dong, Yongheng Hu and Tao Dong, *Member, IEEE*

***Abstract*—Detection of biomarkers in urine sample is often conducted by use of dipsticks, which provides a qualitative result. Urinalysis involving image recognition and data processing has becoming one of the powerful tools in clinical diagnosis. This paper presents colorimetric recognition of urinalysis dipsticks based on quadratic discriminant analysis (QDA) in order to overcome the drawbacks, such as, limited detection area, seriously affected by the external light conditions etc. It can decrease the error of color space conversion by directly processing the data from the captured image using QDA. The correlation of the sRGB color space and the difference of covariance matrix of the acquired data were took into account in this discriminant analysis. The results of validation experiments by Matlab simulation show that it can effectively identify the similarity between the test and reference color on the dipsticks with the color recognition accuracy at 97.33%.**

I. Introduction

Many diseases such as hypertension (HT), dyslipidemia, diabetes mellitus (DM), etc. have becoming the health problems human concerning recent decades [1-2]. These diseases could detected by urinalysis dipsticks [3-4]. The urinalysis by use of commercial dipsticks has to involve the process of urine sample collection, which are time-consuming and even painful especially for the aged. Moreover, the current urine detection methods explored their obvious shortages: instrument recognition with high-cost and time-consuming; color recognition by naked-eye are low efficiency, high error. Therefore, a fast, smart, economical and convenient urinalysis method has becoming an urgent and promising issue. The increasing number of researchers are focusing on that.

Du et al [5] did the color recognition by converting the captured image to YUV color space by procedures of filtering, noise reduction and image segmentation based on the color area of the dipstick to analysis urine components. However, the big difference between final obtained color and standard reference color due to some objective conditions of instrument precision, illumination, etc. is not expected in practical applications. Zhang et al [6] proposed a color signal disease diagnosis method based on lateral-flow strips. The combined pictures, consisting of several pictures taken several times for the same strip, processed with the procedures of filtering, gray-scale transformation, edge detection etc. to get final valid information. This method could infer diseases by quantitative analysis for the test paper (HCG concentration obtained from average grayscale value). Unfortunately, high-cost of hardware equipment, special requirement of taking several pictures at fixed position, complex pre-processing procedures before obtaining grayscale image with single channel (more easily affected by light resource and exposure intensity) make it more different to be popular. Ding et al [7] got human PH values using image processing in sRGB color space: cyclically calculate the Euclidean distance between the test and reference points. The final test PH value is corresponding to reference PH value with minimum distance. It is theoretically reasonable, but the missing of sRGB color correlation makes it to be low accuracy using Euclidean distance.

All these methods with high-cost or low accuracy makes it difficultly popular used in medical diagnosis. Current smartphone pixel is able to satisfy the basic requirements of image recognition. The smartphone-based robust correction algorithm proposed by Karlsen et al [8] for the study of urine analysis demonstrated its possibility of point-of-care disease diagnosis. In this paper, the QDA-based colorimetric recognition of urinalysis dipstick will be studied, that combine with smartphone application as Haakon [8] described. The validation experiments with Matlab simulation reveal its high recognition rate. Moreover, apparent error rate (APER) evaluation of QDA is also presented.

II. Modeling

A. Physical Layout of the Dipstick

The urinalysis dipstick studied in this paper were proposed by Karlsen et al [8]. Here, only Specific Gravity (SG) part of whole multi-biomarker dipsticks analyzed, as shown in figure 1. It include two parts: the reference color blocks with squares locating at the left side of the black triangle, another side of that is the rectangular test block. The test block integrated with dry chemical reagents will change color once reacted with urine sample. Capture picture using smartphone camera after chemical reactions for further analysis.

Figure 1. Standard urinalysis dipsticks (partial only with SG part) [8].

Actually, the square and rectangular blocks in figure 1 are filled with color pixels. The black or gray color presented as simplification.

*This work is supported by Oslofjordfond projects "Touchsensor for enklere og raskere urinprøvetaking og analyse" (no.234972), " Papirbasert kolorimetrisk sensorsystem med integrert polymer-lyskilder og -detektorer for kvantitativ deteksjon av biomarkører i spyt" (no.249017).

Tao Dong is with Department of Microsystems-IMS, Faculty of Technology, Natural Sciences and Maritime Sciences, University College of Southeast Norway, Postboks 235, 3603 Kongsberg, Norway(corresponding author to provide phone: +47-3100-9321; fax: +47 33031103; e-mail: Tao.Dong@usn.no)

Yong He and Kai Dong is with the School of Mechanical Engineering, Nanjing University of Science & Technology (NJUST), contributed equally and shared the first authorship. Yongheng Hu is with the College of Communication Engineering, Chengdu University of Information Technology.

B. The Principle of Image Recognition

Take the images using smartphone after the chemical reactions in the test block (pad). Procedures of lens distortion correction, filtering and smooth processing utilized for the image processing [5]. All these procedures processed in sRGB color space without space conversion because of: the data loss from each color space conversion will affect the recognition accuracy; calculation speed will decrease caused by increased calculation workload due to color space conversion; QDA could eliminate the errors associated with the correlation of sRGB color space.

Select 51*51 pixels on each block to obtain data of reference and testing color after pre-processing of captured image. An *M*×3 matrix created using the pixels data (3 channels (RGB) for each pixel). Then it is easy to get mean vector μ, covariance matrix of the reference point Σ (obtained from sampled data of RGB channels). If the covariance matrixes of reference and testing points are different, QDA are useful for the judgment. Insert testing data into QDA equation to calculate discriminant coefficient of the pixel and obtain mean discriminant coefficient for the test color block. Meanwhile, calculate the mean discriminant coefficient of each reference color block in the same way. The reference color block with the maximum discriminant coefficient is the closest to the test color block.

C. The Procedures of Image Recognition

Figure 2 shows the flow chart of image recognition [9-11].

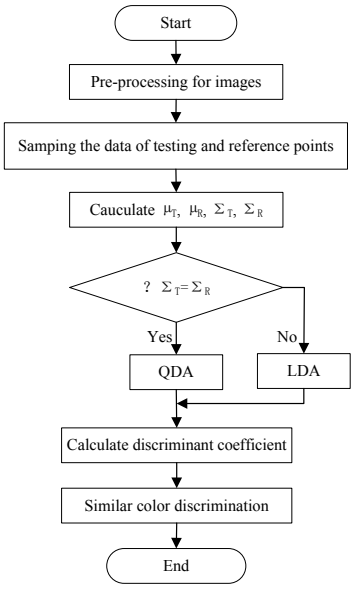


Figure 2. Flow chart of image recognition.

III. Principles and Evaluation of QDA

Bayesian discriminant principle is based on the mastered sample information of each category to establish a discriminant function according to the summarized classification regularity [12]. This function determine that whether pixels/points are correctly classified or not.

A. Principle of QDA

QDA could derived from the conditional probability model [13-14]. Assume that there are *M* categories with prior probabilities of $P(C_1), P(C_2), \cdots P(C_M)$. A random sample selected with $X(x_1, x_{2,} \cdots x_p)$, x_p stands for p category (here $p = 3$). We classify as C_i if it satisfies the equation (1):

$$P(C_i/X) > P(C_j/X), \quad j \in 1, 2, \cdots M. \tag{1}$$

Bayesian discriminant principle:

$$P(C_i/X) = \frac{P(C_i)P(X/C_i)}{P(X)}. \tag{2}$$

QDA is based on multivariate Gaussian distribution model, and the Gaussian probability density function shown in equation (3) [14].

$$f_i(x) = \frac{P(C_i)P(X/C_i)}{(2\pi)^{p/2}|\Sigma_i|^{1/2}} exp\left[-\frac{1}{2}(x-\mu_i)'\Sigma_i^{-1}(x-\mu_i)\right]. \quad i \in 1, 2, \cdots M \tag{3}$$

Seeking the logarithm for equation (3) to deviate the QDA formula:

$$d_i^Q(x) = -\frac{1}{2}ln|\Sigma_i| - \frac{1}{2}x'\Sigma_i^{-1}x + \mu_i'\Sigma_i^{-1}x - \frac{1}{2}\mu_i'\Sigma_i^{-1}\mu_i + lnP_i. \quad i \in 1, 2, \cdots M \tag{4}$$

From equation (4), it is easy to get that QDA discriminant coefficient consists of: 1) Generalized variance $|\Sigma_i|$; 2) Prior probability $P_i = 1/N$, each reference color has the same probability chosen in *N* reference color; 3) Mean vector μ_i, covariance matrix Σ_i:

$$\mu_i = \frac{P(C_i)P(X/C_i)}{P(X)}. \tag{5}$$

$$\Sigma_i = \begin{pmatrix} cov(x_{:,1}, x_{:,1}) & \cdots & cov(x_{:,1}, x_{:,p}) \\ \vdots & \ddots & \vdots \\ cov(x_{:,p}, x_{:,1}) & \cdots & cov(x_{:,p}, x_{:,p}) \end{pmatrix}. \tag{6}$$

51*51 pixels selected at testing area as the sample. Insert the testing data to equation (4) to calculate the average discriminant coefficient after obtaining the corresponding mean vector, covariance matrix and the inverse covariance matrix of the reference blocks. The same procedure as each reference color block. Compare average discriminant coefficient of each reference block to find the maximum, which is the most closet to the test block/pad.

B. Evaluation of Discriminant analysis

Error rate is the important criteria to evaluate whether the performance of discriminant analysis is superior enough or not. Apparent error rate (APER), as a discriminant method of normally used error rate, applies to any distribution type and defined as the study of erroneous classification of training samples [15-17].It needs to satisfy the requirements of: 1) Enough training samples; 2) All the samples must participate in the construction of the classifier in case of losing of useful information. In other words, APER can be described as:

$$APER = \frac{n_1 m + n_2 m}{n_1 + n_2}. \tag{7}$$

TABLE I. Predict Variable Relationship for APER

	A	***B***
Actual number of A	$n_1 c$	$n_1 m = n_1 - n_1 c$
Actual number of B	$n_2 m = n_2 - n_{21} c$	$n_2 c$

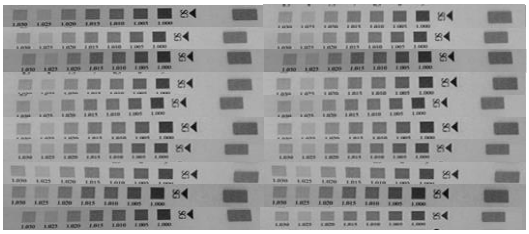
n_1c: samples correctly classified to A; n_1m: samples misclassified to B; n_2m: samples misclassified to A; n_2c: samples correctly classified to B.

IV. Experiments and Results

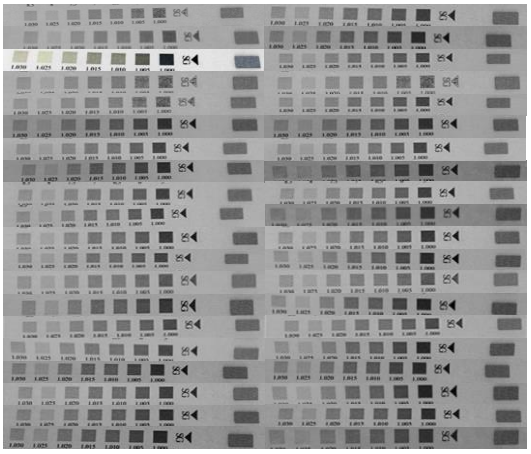
A. Experimental Procedures

In order to verify the validity of this QDA method, four group urinalysis dipsticks selected for recognition analysis under different lighting conditions, as shown in figure 3 (A with 5 dipsticks, B with 10 dipsticks, C with 20 dipsticks and D with 40 samples, respectively. Only group C and D presented in figure 3). SG as the test indicator.

Experimental steps: (1) 51*51 pixels obtained from each reference and test color block; (2) Calculate the mean vector and covariance matrix of reference points. Determine whether the covariance matrix is equal and find the inverse matrix; (3) Insert the reference data into



Group C



Group D

Figure 3. Experimental captured pictures for group C and D.

equation (4) to obtain average discriminant coefficient of test color block and each corresponding reference color block; (4) Get the serial number of reference color block with the largest average discriminant coefficient; (5) Obtain the result of that the test color block is the most closet to the reference color block with serial number obtained in last step.

B. Simulation with Matlab

Here, in order to verify the validity of this discriminant method, some validation experiments with Matlab simulation are necessary. Furthermore, it is also necessary to do the prior settings of that the testing color block is the most closest to the specified reference color block (group A to D are corresponding to 2, 3, 5, 6, respectively). Do the

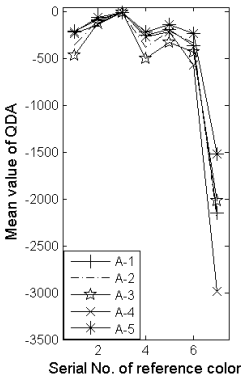


Figure 4. QDA discriminant results of group A.

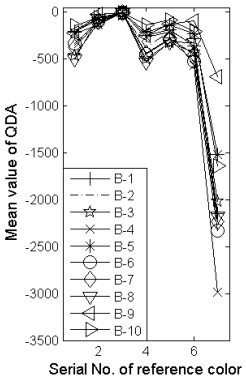


Figure 5. QDA discriminant results of group B.

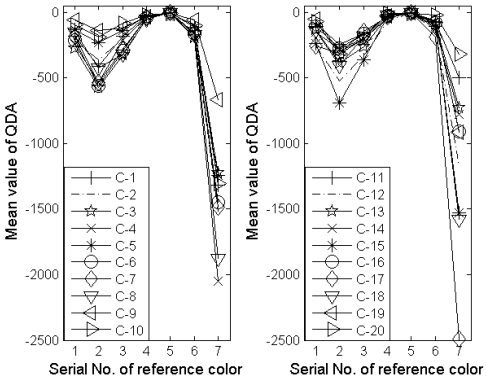
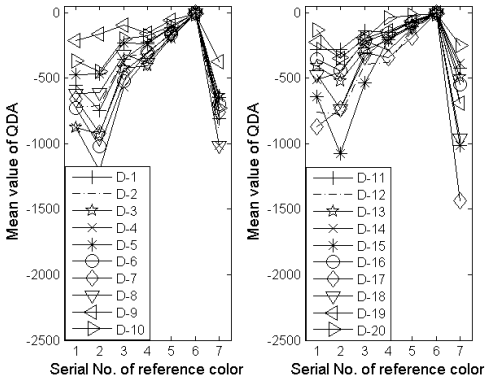
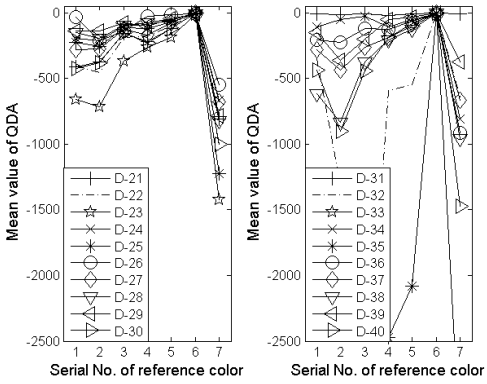


Figure 6. QDA discriminant results of group C.



(a)



(b)

Figure 7. QDA discriminant results of group D.

simulations according to the experimental steps above with Matlab for group A to D.

The simulation results shown in figure 4 to 7 indicate that it can obtain the maximum at the points of polyline for the reference color block (group A at No. 2, group B at No. 3, group C at No. 5 and group D at No. 6 respectively), which are coincident with the prior settings of serial number above. There are two curves occur the recognition error (wrong recognition), but the total color recognition rate can reach as high as 97.33%.

C. APER Evaluation

APER is an important indicator of the robustness and superiority of the algorithm. The correct classification of each pixel of the testing color is the key point that whether the final color recognition is correct. Build the APER model: (1) Calculate APER of each pixel of sampled testing and reference color. (2) Obtain minimum APER by comparison. (3) If the minimum APER is the same as pre-setting reference color (Here, group A with 2, group B with 3, group C with 5 and group D with 6, respectively), the classification is correct. Otherwise, it is wrong.

TABLE II. APER ANALYSIS RESULTS

Group	*2601 pixels (51*51)*		*Discriminant rate (%)*	*APER (%)*
	Average correct recognition pixels	*Average wrong recognition pixels*		
A	2028.78	572.22	78.00	22.00
B	2282.38	318.26	87.75	12.25
C	2244.66	356.34	86.30	13.70
D	2220.21	380.79	85.36	14.64

APER analysis results, 51*51 pixels obtained from each testing color, as following in table II. Table II shows that mostly 51*51 pixels for each group could be correctly discriminated (average discrimination rate: 84.35%), though influenced by the lighting conditions.

V. CONCLUSION

The QDA-based method for the study of colorimetric recognition in medical urinalysis dipsticks is presented in this paper, which can make a correct discrimination (recognition rate can reach as high as 97.33%) by validation experiments with Matlab simulation. In addition, the average correctly discrimination rate (classification) for the pixels is 84.35%. The results show that the undeniable superior of QDA in colorimetric recognition of urinalysis dipsticks. It can overcome the negative effects of small sample size, external environment conditions etc. That is meaningful and valuable for the study of colorimetric recognition for urinalysis dipsticks based on QDA.

However, image with JPEG format will increase the correlation of data analysis for small samples, which will affect the final discriminant analysis. Hence, it is necessary to do the comparison between QDA and linear discriminant analysis (LDA) for this colorimetric recognition in the following research work.

ACKNOWLEDGMENT

The authors acknowledge the "Smart-tøy for eldreomsorg: Oppfølging av fysiske aktiviteter og overvåkning av fysiologisk status i sanntid" (no.260586), and Forprosjekt – VRIBEDRIFT Vestfold "mikrofluidikkbasert PCR med kontinuerlig strømning i sanntid, for rask deteksjon av mikrobielle patogener i vann" (no. 252173). This research is also supported by National Natural Science Foundation of China (no. 61550110253, 61650410655), Chongqing Research Program of Basic Research and Frontier Technology (nos. cstc2016jcyjA2161, cstc2015jcyjA20023), Science and Technology Research Program of Chongqing Education Commission (no. KJ1600604) and Chongqing Innovation Team of Colleges and Universities (no. CXTDX201601025).

REFERENCES

[1] Maahs D M, Eckel R H. Type 1 Diabetes Mellitus and Dyslipidemia. Dyslipidemias, 2015, pp.115-135.
[2] Hall J E. Hypertension—Opportunities and Challenges. Hypertension, 2002, 39(1), pp.1-2.
[3] Krogsbøll L T, Jørgensen K J, Gøtzsche P C. Screening with urinary dipsticks for reducing morbidity and mortality. Cochrane Database of Systematic Reviews, 2015, 1, pp.CD010007.
[4] Zolghadri J, Alborzi S, Ardekany M. THE VALUE OF DIPSTICK ANALYSIS OF URINARY PROTEIN IN PREGNANCY INDUCED HYPERTENSION. Medical Journal of the Islamic Republic of Iran, 2002.
[5] Du Y.X., Xu J.H., Application of Image Processing Technology in Urine Testing. Microcomputer information, 2011, 27(1), pp.250-252.
[6] Zhang X, Yang H, Wang K. Image processing for the CCD based lateral flow strip detector. Nano Biomedicine & Engineering, 2010, 2(4).
[7] Ding Y, Wang Q. A rapid quantitative detection method of colorimetric test strip based on machine vision. 2011, 9(3), pp.836-839.
[8] Karlsen H., Dong T., A Smart Phone-Based Robist Correctopm Algorithm for the Colorimetric Detection of Urinary Tract Infection", 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2015.
[9] Image recognition. Photonics Spectra, 2006.
[10] Höppner F, Klawonn F, Kruse R, et al. Fuzzy Cluster Analysis: Methods for Classication, Data Analysis and Image Recognition. Journal of the Operational Research Society, 2000, 51(6), pp.769-770.
[11] Wang J, Wu X, Lu Y, et al. Face recognition in simulated prosthetic vision: face detection-based image processing strategies. Journal of Neural Engineering, 2014, 11(4), pp.046009.
[12] https://en.wikipedia.org/wiki/Bayesian_inference
[13] Curtis A. Classification using LDA, QDA and logistic regression. 2005, 3.
[14] Laurinda F.S. Siqueira, Aurigena Antunes de Araújo, Camilo L.M. Morais, et al. LDA vs. QDA for FT-MIR prostate cancer tissue classification. Chemometrics and Intelligent Laboratory Systems, 2017:123–129.
[15] Xanthopoulos P, Pardalos P M, Trafalis T B. Linear Discriminant Analysis. Robust Data Mining. Springer New York, 2013:2464-2485.
[16] Zhang J, Shen H, Zhou Z H. Unified Locally Linear Embedding and Linear Discriminant Analysis Algorithm (ULLELDA) for Face Recognition. Advances in Biometric Person Authentication. DBLP, 1970:105-120.
[17] Bradley Efron. How Biased is the Apparent Error Rate of a Prediction Rule? Journal of the American Statistical Association, 1986, 81(394), pp.461-470.