Title	Informative Census Transform for Very Low-Resolution Image Representation
Author(s)	Jeong, Sungmoon; Lee, Hosun; Chong, Nak Young
Citation	IEEE International Symposium on Robotand Human Interactive Communication: 435-440
Issue Date	2014-08
Type	Conference Paper
Text version	author
URL	http://hdl.handle.net/10119/12344
Rights	This is the author's version of the work. Copyright (C) 2014 IEEE. IEEE International Symposium on Robotand Human Interactive Communication, 2014, 435-440. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.
Description	

Informative Census Transform for Very Low-Resolution Image Representation

Sungmoon Jeong, Hosun Lee, and Nak Young Chong

Abstract— **Our paper newly presents unsupervised feature representation method for very low-resolution (VLR) images called informative census transform (ICT) based on statistical analysis of CT binary features and submodular optimization. A new cost function is designed to measure the informativeness of each binary feature: (1) an individual informativeness of features to represent unlabeled image dataset and (2) relative informativeness between binary features to represent different binary features. Therefore, we considered informativeness of binary feature according to two relationship (1) between feature space and image space, and (2) between different features within same feature space. Moreover, two constraints are designed by considering sub-modular characteristics to guarantee theoretical performance and fast optimization via simple greedy algorithm. Experimental results show that the proposed ICT features with two constraints outperforms the traditional CT features in terms of recognition performance and computational cost at VLR problem.**

I. Introduction

Visual feature extraction and representation method is essential task to have a good classification and recognition performance in vision based robot application. Various feature representation methods based on shape information have been widely proposed to recognize and localize the object which are briefly divided by two categories: local descriptor based image representation [1]-[3], and global feature based representation [4].

However, in case of mobile robotics or surveillance system, the less image resolution we have, correlation coefficient between different images is increased and it causes poor discriminant characteristics of feature extraction and representation methods. Especially in very low resolution (VLR) recognition problem, such as CCTV and small mobile robot, which has less than 16 by12 pixel size, above approaches could not be directly applied for VLR problem because most of feature extraction methods require enough image resolution (64 by 48 image size) [5]. Moreover, coverage area of each RGB sensor and pixel has much broad field of view to sense a physical environment that causes redundant features of VLR images even though they have different spatial characteristics.

Even each pixel covers huge physical area, spatial structure of neighbor pixels is very important to extract valuable features at VLR problem. Among the feature representation methods, local binary pattern (LBP) and modified census transform (MCT) show good classification performances as well as low computational cost in real world application because they rely on the local intensity order is unaffected by linear and monotonic transformation on the reflectance [6]. Based on this characteristic, these models robustly analyze spatial structure of an image based on binary patterns and this spatial information is used to classify visual images by comparing the similarity between each image based on hamming distance. However, in VLR recognition problem, many of these binary features might be uninformative or redundant even these approaches are based on structure analysis. It means that each binary code bit has different effects on the representation of visual images. For instance, some of binary features are almost the same value for entire training images and these redundant features cause poor discriminability and much computational cost. Therefore, the selection of informative-structured feature is required to efficiently represent an image.

Recently, the property of submodularity, guarantees theoretical performance and fast optimization by using simple greedy algorithm, is investigated and examined on several domains [7], [8]. Based on this property, high dimensional feature selection was presented by [9] using the uncapacitated facility location function and the saturated coverage function. However, two cost functions are designed by only considering relative informativeness between features within the same feature space to select features which can effectively represent different binary features. If the number of selected features is increased and/or in such case of low dimensional feature selection problem, correlation coefficient between different features is also increased and discriminant characteristics of selected features are decreased. Thus, individual informativeness of each feature is important factor for feature selection to represent an image. In other words, relationship between image space and feature space is needed to measure the informativeness of each feature.

To solve the VLR recognition problem, existing methods considered image reconstruction method from VLR to high resolution (HR) image based on relationship learning between pairs of LR and HR images [10]-[12]. However, there are two requirements to apply existing approaches: (1) HR reference images to learn about relationship between LR and HR and (2) image labels as supervised manners. But, it is difficult to satisfy the aforementioned requirements at most of real application problem.

Our focus lies on informative-structured feature extraction and selection method with unsupervised manners to enhance discriminability and computational efficiency in VLR recognition problem. We newly propose an informative census transform (ICT) based on the CT and statistical analysis with unlabeled training data. To select informative CT feature, a cost function is designed to simultaneously

S. Jeong, H. Lee, and N. Y. Chong are with the School of Information Science, Japan Advanced Institute of Science and Technology, Ishikawa, Japan (corresponding author to provide phone: 81-761-51-1362; e-mail: jeongsm@jaist.ac.jp).

optimize two informativess: (1) an individual feature informativeness based on entropy information [13] to effectively represent training set (relationship between training image space and feature space) and (2) relative feature informativeness using mutual information between features [9] to efficiently represent different features (relationship between different features within same feature space). Moreover, to guarantee the optimal solution, the cost function is designed to satisfy submodularity, and finally near optimal solution is efficiently achieved with simple greedy algorithm. We have shown that the proposed ICT feature with two constraints outperforms conventional CT features in VLR recognition problem.

The novelty of this paper is as follows: (1) integration of two submodular cost functions to cover both characteristics of feature space and image space have never before been used for binary feature selection in VLR image recognition problem. (2) Moreover, unsupervised image representation approach with only VLR image is novel to recognize VLR image while existing methods require HR images and image labels.

The remainder of this paper is organized as follow: Section II describes the proposed models for informative census transform. Experimental results and analysis are presented in Section III, and conclusions and a discussion of further work are given in Section IV.

II. METHODS

A. Census transform to extract local binary structure

The Census Transform (CT) is a non-parametric local structure extraction which was first proposed by Zabih and Woodfill [6]. It is defined as an ordered set of comparisons of pixel intensities in a local neighborhood representing which pixels have lesser intensity than the center. Therefore, center pixel is not contained to represent local spatial structure. The MCT used average value within local spatial structures instead of center pixel to obtain a result for the center pixel. Within the kernel structure information is coded as binary information and the resulting binary patterns can represent oriented edges, line segments, junctions, ridges, saddle points, etc [3]. Local spatial structure $\Gamma(x_i)$ of the ith pixel at x_i is calculated as

$$\Gamma(x_i)_{x_i \in X} = \otimes_{y \in Nx_i} \xi\big(\bar{I}(x_i), I(y)\big) \quad (1)$$

where $\otimes$ denotes the concatenation operation, *Nx* is the neighborhood pixels within x position, $\bar{I}(x)$ is mean intensity on the neighborhood pixels, and $\xi\big(\bar{I}(x), I(y)\big)$ be 1 if $\bar{I}(x) < I(y)$ and otherwise 0. Based on this concatenated binary feature, we can encode from binary to decimal value to reconstruct visual images. Fig. 1 shows the examples of transformed image by MCT binary structure features. As shown in Fig 1, transformed face images have more robust characteristics to represent face images by reducing the illumination effects. Finally, each local binary structure feature $\Gamma(x_i)$ are concatenated to represent an image X and this binary feature set ($\mathrm{BFS} = \otimes_{i=1,\dots,Nx}\Gamma(x_i)$ is compared with different binary feature to classify whether same image class or not by calculating hamming distance. Therefore, each binary feature has same effect to calculate the hamming distance. However, according to the image classes, each binary feature has different representation ability to increase the discriminant characteristics between different image classes. In such case of VLR image recognition problem, lots of features are redundant to represent different image classes.

(a) (b)

Figure 1. Examples of transformed image using MCT binary structure features. (a) original images, (b) transformed images.

B. Informative Census Transform (ICT)

Submodular functions are a class of discrete functions that have "*diminishing returns*" property which means that a function *f* is called submodular iff, for all subsets X, Y belong to a finite set $\mathbb{S}$ $\left(^{\forall}\mathrm{X}, {}^{\forall}\mathrm{Y} \subset \mathbb{S}\right)$ and all singletons $\mathrm{s} \in \mathbb{S}$, we have $\mathrm{X} \subset \mathrm{Y} \Rightarrow f(\mathrm{X} \cup \{\mathrm{s}\}) - f(X) \geq f(\mathrm{Y} \cup \{\mathrm{s}\}) - f(Y)$. Among the submodular functions, a function f is called monotone submodular function iff, for $^{\forall}\mathrm{X} \subset \mathbb{S}$, we have $f(\mathrm{X} \cup \{\mathrm{s}\}) - f(X) \geq 0$. Then, the problem of maximizing $f(X)$ can be nearly optimized by greedy algorithm and performance $f_{greedy}(X)$ is guaranteed larger than $(1 - 1/\mathrm{e}) f_{optimal}(X)$ [7].

According to this property, we design the cost function for binary structure feature selection problem. There are two main constraints to design the cost function for binary feature selection, (1) inner-informativeness within each binary feature to efficiently represent training image set, (2) intra-informativeness between each binary feature to effectively represent different features. Two constraints are formulated by considering monotone submodular function and combined to select most informative feature set of *s*.

Individual informativeness of binary feature

In information theory, information entropy is a measure of the uncertainty. Moreover, information entropy is the average unpredictability in a random variable, which is equivalent to its information content. It means that informative contents have much information entropy value than useless information contents. Therefore, informative contents are useful to represent to different images. Therefore, an entropy value of each binary feature is used to measure the individual informativeness of a feature and calculated by

$$\mathrm{H}\big(\mathrm{BF}_j\big) = -\sum_{k=0}^{K=1} p\big(\mathrm{BF}_{jk}\big) log p\big(\mathrm{BF}_{jk}\big) \quad (2)$$

where BF_j denotes jth binary feature and $p\big(\mathrm{BF}_{jk}\big)$ is probability distribution of binary value k at jth binary feature using unlabeled training set. Each entropy value is calculated by probability distribution of binary features, respectively.

Relative informativeness between binary feature

Recently, the uncapacitated facility location function was used to measure a relative informativeness between high dimensional features to select informative features [9]. This function was successfully applied to select feature selection to represent different features. In order to apply this function to select binary features, we considered probability distribution of binary features as new feature vector for each binary digit and measured the representative capability score to present different features as

$$\mathrm{U}(\mathrm{BF}_S) = \sum_{j=1}^{J} \max_{s \in S} exp - \left\| p\left(\mathrm{BF}_j\right) - p(\mathrm{BF}_s) \right\| \qquad (3)$$

where BF_S denotes selected binary feature set S among the entire J binary feature set.

Finally, we designed monotone submodular function to select informative binary feature by weighted combining Eqs. (2) and (3) as

$$\underset{S}{\operatorname{argmax}} \left\{ \alpha \sum_{s=1}^{S} \mathrm{H}(\mathrm{BF}_s) + (1-\alpha)(\mathrm{U}(\mathrm{BF}_S)) \right\} \qquad (4)$$
$$0 \leq \alpha \leq 1$$

where α represents integration ratio between individual and relative informativeness for informative binary feature selection to efficiently represent not only for different features but also training image dataset. The integration ratio is manually set in current work and will be optimized by several optimization techniques such as evolutionary computation.

III. Experimental results

A. Datasets

To demonstrate the effectiveness of the proposed model, the AR face database [14] and Extended Yale B face database [15] are used. Moreover, two classifiers, such as Naïve Bayes (NB) classifier, and sparse coding with L2 minimization using QR decomposition (L2) [16], are used for both of the proposed ICT feature and traditional MCT feature with 3x3 kernel structure. The face dataset was encoded by binary structure values and all binary features are concatenating as one feature vector to represent an image.

AR face database

The AR database consists of over 4,000 frontal images for 126 subjects [15]. We randomly select a subset of the data set consisting of 50 male subjects and 50 female subjects. For each subject, 26 pictures consist of 14 non-occlusion faces and 12 occlusion faces. We considered only normal faces and half of them are used to train feature representation method and rest of faces are evaluated to calculate recognition accuracy. These non-occluded images include facial variations, illumination change. In the experiment, the images are cropped and converted to gray scale with 165 by 120 and resized with 10 by 8 image size as shown in Fig. 2.

Extended Yale B face database

The Extended Yale B database consists of 2,414 frontal-face images with different illumination for 38 individual subjects [14]. We randomly split same number of training images and test images as 25 for each subject. The images were cropped with 192 by 168 image size and resized by 10 by 8 image size as shown in Fig. 3. Finally, VLR images with 10 by 8 image size of two database are encoded by MCT features with 720 dimension and a part of features among the 720 features is selected by maximizing the three constraint such as (1) individual informativeness as entropy information (CS1), (2) relative informativeness as uncapacititated facility location function (CS2) and (3) their integration (CS3).

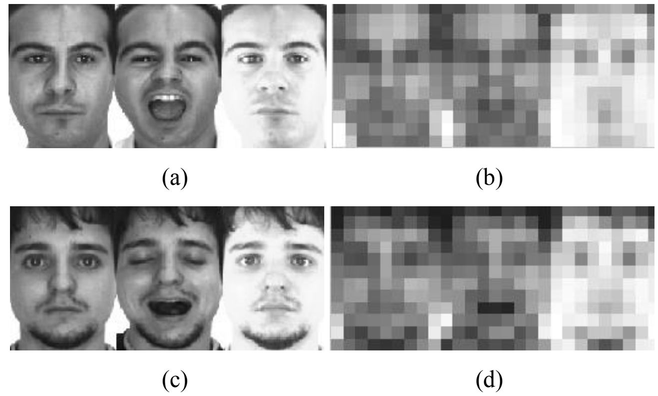


Figure 2. AR face dataset with HR and LR. (a), (c) examples of AR face dataset with HR image considered facial variations, illumination change (165 by 120 image size). (b), (d) LR face images of (a) and (c) (10 by 8 image size).

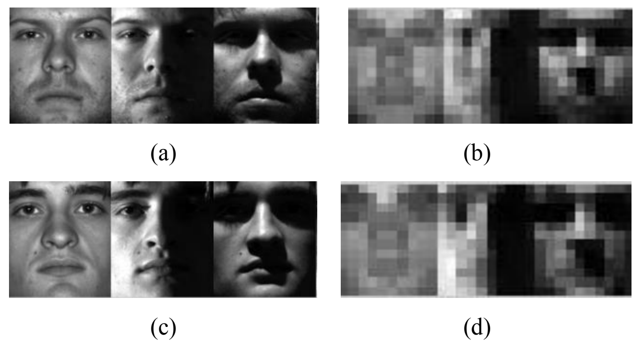


Figure 3. Extended Yale face dataset with HR and LR. (a), (c) examples of extended Yale face dataset with HR image considred different illumination conditions (192 by 168 image size). (b), (d) LR face images of (a) and (c) (10 by 8 image size).

Discriminant characteristics according to different image resolution

Figs. 4 and 5 shows the discriminant characteristics between different images according to image resolution. The discriminant characteristics are calculated by average Euclidean distance between face images. As shown in Figs. 4 and 5, discriminant characteristic is decreased by increasing down sampling rate.

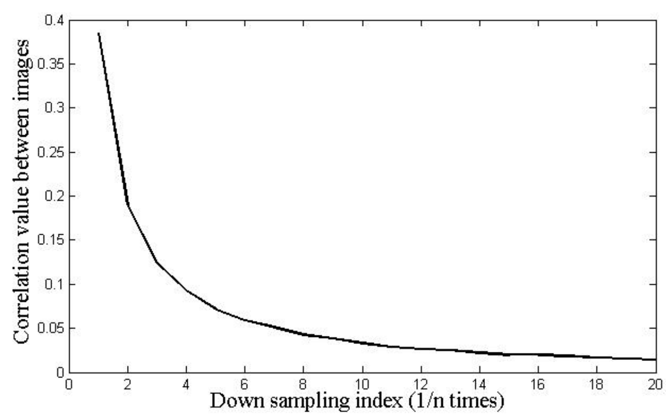


Figure 4. Relationship between discriminant characteristic of different images and image resolution using AR face dataset (from 165 by 120 image size to 8 by 6 image size).

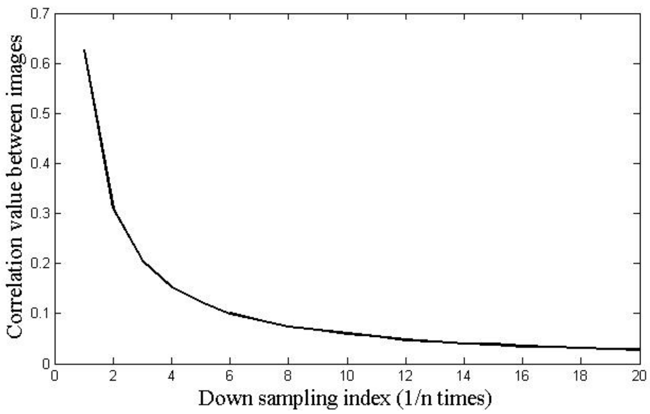
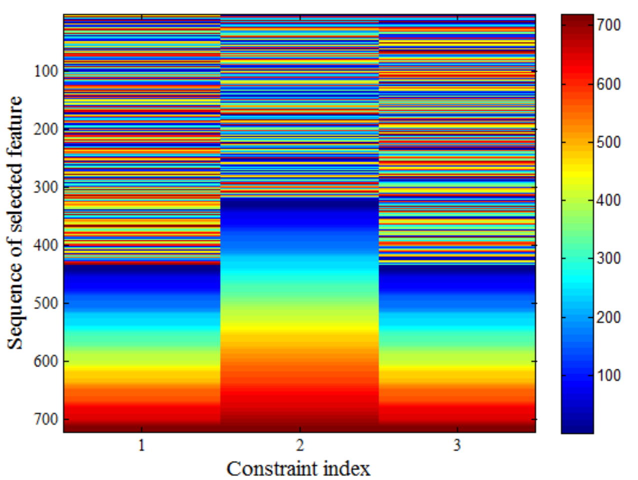


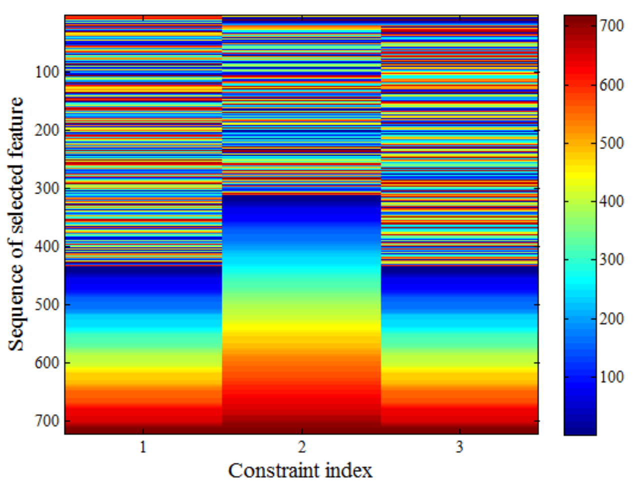
Figure 5. Relationship between discriminant characteristic of different images and image resolution using extended YaleB face dataset (from 192 by 168 image size to 9 by 8 image size).

B. Comparison between ICT features with different constraints

Fig. 6 shows the sequence of selected features with different face database using three constraint.



(a)



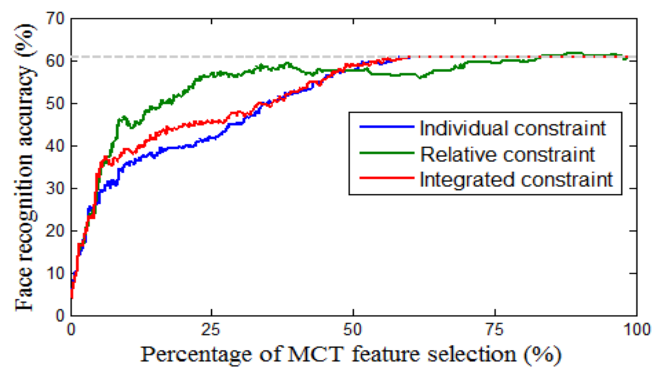
(b)

Figure 6. Sequences of selected binary features by maximizing three constraints. (a) Extended Yale B face dataset: after 60 % of feautres (around 400 features) are selected then remained 40% of features (around 300 features) are sequentially selected by maximizing CS1 and CS3 because they have same characteristic. On the other hand, remained 60% of features are sequentially selected by considering CS2. (b) AR face dataset. Observed results for AR face datset are similar with Yale B face dataset (a).

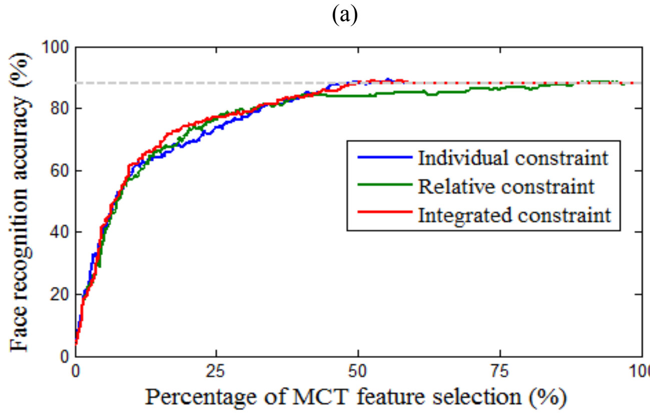
In case of VLR image recognition problem, lots of features have redundant characteristics. In case of relative informativeness function (CS2) for Yale B database as middle column in Fig. 6 (a), after 40 % of features around 300 features are optimally selected and then remained 60% of features around 400 features are sequentially selected as shown in Fig. 6 (a) because remained features have same informativeness value using relative constraint. On the other hand, remained around 40% of features are sequentially selected by maximizing individual constraint (CS1) and integrated constraint (CS3) as shown in Figs. 6 (a) and (b). It means that individual informativeness value can be used to increase the discriminant characteristics of binary features in VLR image recognition problem.

C. Face recognition results

Figs. 7 and 8 show the classification results for extended Yale B face dataset and AR face database according to the number of selected features and Table I and II represent maximum recognition accuracy based on different feature types. We can observe that three constraints are useful to select informative binary feature for face recognition problem.



(a)



(b)

Figure 7. Recognition rates for extended Yale B face database using ICT features with different constraints. Gray dashed line: traditional MCT (baseline method), blue line: individual informativeness (entropy), green line: relative informativeness (uncapacititated facility location function), red line: integration of two

informativeness. (a) Naïve Bayesian (NB), (b) Sparse coding using L2 minimization and QR decomposition (L2).

Especially, the integrated constraint has better or similar recognition performance and less number of features than other constraints. Moreover, in case of relative constraints as green line in Figs. 7 and 8, remained 60% of features are not useful to enhance the face recognition accuracy because of large correlation coefficient between different features as shown in Figs. 7 and 8. On the other hand, individual informativeness helps to increase the face recognition performance by additionally selecting 20% of features.

TABLE I. COMPARISON OF CLASSIFICATION PERFORMANCE USING YALE B FACE DATASET WITH DIFFERENT CONSTRAINTS AND CLASSIFIERS

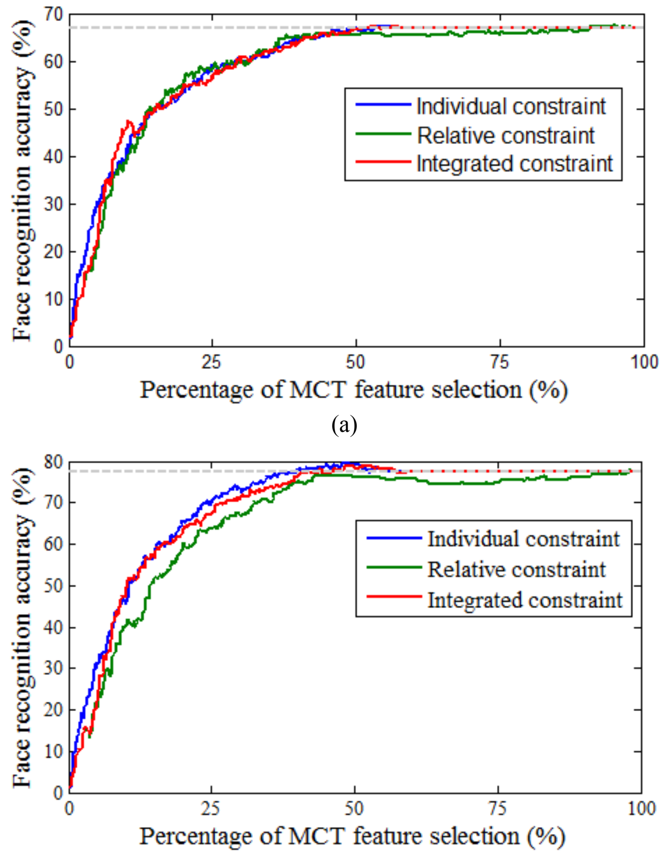
Feature and classifier types	Number of selected features for maximum accuracy	Classification accuracy (%)
MCT-NB	720	60.6
CS1-NB	418	60.8
CS2-NB	641	61.8
CS1&2-NB	430	60.8
MCT-L2	720	88.1
CS1-L2	397	89.3
CS2-L2	682	88.5
CS1&2-L2	410	89.2

TABLE II. COMPARISON OF CLASSIFICATION PERFORMANCE USING AR FACE DATASET WITH DIFFERENT CONSTRAINTS AND CLASSIFIERS

Feature types	Number of selected features for maximum accuracy	Classification accuracy (%)
MCT-NB	720	66.9
CS1-NB	406	67.4
CS2-NB	683	67.6
CS1&2-NB	382	67.4
MCT-L2	720	77.4
CS1-L2	347	79.9
CS2-L2	707	78.0
CS1&2-L2	350	79.0

IV. CONCLUSION

In this paper, we proposed new feature representation method called informative census transform (ICT) for VLR image recognition problem only using VLR images. To optimally select informative binary features, cost function is designed by integrating individual and relative informativeness of binary structures according to the property of submodular functions. Experimental results show that proposed ICT feature with three constraints have better classification performance and less number of selected features than traditional MCT feature and also integrated constraint are more useful to select informative features than others.



(b)

Figure 8. Face recognition rates for AR face database using ICT features with different constraints. The figure legends are the same as for Fig. 7. (a) Naïve Bayesian, (b) L2 minimization using QR decomposition.

Therefore, in such case of VLR problem, both of individual and relative informativeness are important factors to solve the informative feature selection task. The proposed algorithm can be applied to face recognition for human robot interaction system with very low resolution camera and various recognition and classification application in robot vision research field.

As a further work, various VLR image database and existing methods will be considered to verify the proposed algorithms and analyze the specific role of each constraint. And, we will investigate how to integrate two constraints by applying optimization techniques. Also, by applying adaptive submodular functions, adaptive ICT features will be considered to be applied in real time application problem with VLR image such as dynamic environment modeling in CCTV camera, stereo vision in VLR cameras and human tracking system using VLR video.

REFERENCES

[1] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” Proceedings on IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), 2005.

[2] T. Ahonen, A. Hadid, M. Pietikainen, “Face recognition with local binary patterns,” Lecture Notes in Computer Sciences, vol. 3021, pp. 469-481, 2004.

[3] B. Fr¨oba, A. Ernst, “Face Detection with the Modified Census Transform,” Proceedings on FGR 04, 2004.

[4] A.M. Martinez, A.C. Kak, “PCA versus LDA,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, no. 2, pp. 228-233, 2001.

[5] J. Wang, C. Zhang, and H.-Y. Shum, “Face image resolution versus face recognition performance based on two global methods,” Proceedings on Asian Conference on Computer Vision, 2004.

[6] R. Zabih and J. Woodfill, “A non-parametric approach to visual correspondence,” Proceedings of ECCV 94, 1994.

[7] A. Krause and C. Guestrin, “Near-optimal Nonmyopic Value of Information in Graphical Models,” Proceedings of Uncertainty in Artificial Intelligence, Edinbugh, Scotland, pp.324-331, 2005.

[8] D. Golovin and A. Krause, “Adaptive Submodularity: Theory and Applications in Active Learning and Stochastic Optimization,” Journal of Artificial Intelligence Research, vol. 42, no. 1, pp.427-486, 2011.

[9] Y. Liu, K. Wei, K. Kirchhoff, Y. Song, J. Bilmes, “Submodular feature selection for high-dimensional acoustic score spaces,” Proceeding of ICASSP 2013, 2013.

[10] W. W. Z. Wilman, and C. Y. Pong, “Very Low Resolution Face Recognition Problem,” IEEE Tansactions on Image Processing, vol. 21, no. 1, 2012.

[11] S. Siena, V. N. Boddeti, B.Y.K. Y. Kumar, “Coupled marginal fisher analysis for low-resolution face recognition,” Proceedings on ECCV 2012, vol. 2, pp. 240-249, 2012.

[12] S.-W. Lee, J. Park, and S.-W. Lee, “Lowresolution face recognition based on support vector data description,” Pattern Recognition, vol. 39, pp. 1809-1812, 2006.

[13] S. Jeong, H. Lee, Y. E. Hamdi, and N. Y. Chong, “Weighted Census Transform for Feature Representation,” in URAI 2013, 2013.

[14] K.-C. Lee, J. Ho, and D.J. Kriegman, “Acquiring linear subspaces for face recognition under variable lighting,” IEEE Trans. Pattern Anal. Mach. Intell. vol. 27, no. 5, pp. 684-698, 2005.

[15] A. M. Martinez, "The AR face database," CVC Technical Report, vol. 24, 1998.

[16] Q. Shi, A. Eriksson, A. van den Hengel and C. Shen, “Is face recognition really a compressive sensing problem?” Proceeding on CVPR 2011, 2011.