Automatic Colorization of Near-Infrared Monochrome Face Image based on Position-Dependent Regression

ATSUSHI MORI[†1] TOSHIKAZU WADA[†1]
HIROSHI OIKE[†1]

This report presents a method for estimating color face images from near-infrared monochrome face images. This estimation is done by the regression from a monochrome image to a color image. One difficult problem is that the regression depends on face organs. That is, the same intensity pixels in an infrared monochrome image do not correspond to the same color pixels. Therefore, entirely uniform regression cannot colorize the pixels correctly. This report presents a colorization method for monochrome face images by position-dependent regressions, where the regression coefficients are obtained in different image regions corresponding to facial organs. Also, we can extend the independent variables by adding texture information around the pixels so as to obtain accurate color images. However, unrestricted extension may cause multi-collinearity problem, which may produce inaccurate results. This report also proposes CCA based dimensionality reduction for avoiding this problem. Comparative experiments on the restoration accuracy demonstrate the superiority of our method.

1. Introduction

Recently, low-cost near-infrared cameras are widely used for security purpose due to its sensitivity to invisible Near-InfraRed (NIR) ray. Such cameras are used with NIR illumination to capture images in nighttime. They are attached to various locations and they sometimes take crime scenes such as Figure 1. On crime investigation, it is effective to estimate a Visible Light Band (VLB) color image from a monochrome NIR image of a suspected criminal.

The NIR face image can be colorized by the regressions between monochrome NIR face images and color VLB face images. The difficulty of this problem is an intensity value in a monochrome NIR image can correspond to multiple RGB color vectors in VLB image. Especially, NIR light is not much absorbed by melanin pigment in iris. This phenomenon produces the intensity order changes in the regression. Therefore, the uniform regression over an entire image cannot correctly colorize NIR face images.

For avoiding this problem, this report presents position-dependent regression, which estimates RGB colors depending not only on corresponding NIR intensities, but also the face organs in which the pixel included.

The independent variable in this regression is the monochrome intensity value. This can be extended to a multi-dimensional vector by adding texture information around the pixel so as to estimate accurate color image. However, unrestricted extension may cause multi-collinearity problem, which may produce inaccurate results.

For avoiding this problem, this report also proposes CCA based dimensionality reduction of independent variables.

In the following sections, we present related works in Section 2 and our method in Section 3. Then we demonstrate the effectiveness through the experiments in Section 4 and conclude the discussion in Section 5.

2. Related works

At the initial stage of colorization research, some methods require the human intervention. Horiuchi and Hirano [1] proposed a colorizing method from several color-assigned pixels in a given monochrome image. The method colorizes the image by propagating the assigned colors from the pixels. The method is computationally inexpensive and can colorize starting from few pixels. However, the method sometimes produces an artifact, i.e., block noise. Levin *et al*. [2] proposed a method that colorizes by propagating colors from user-annotated color scribbles. Irony *et al*. [3] proposed an example-based colorization that performs color transfer over segmented image regions and voting for guaranteeing the spatial consistency. Liu *et al*. [4] also proposed an example-based color transfer in illumination independent domain. The method automatically produces the color scribbles and performs color propagations. This method has an advantage that it does not require image segmentation. These methods are designed for colorizing general scenes, and hence, they require example images obtained by Web search engine.

For face images, colorization is not widely studied. Instead of this, estimating VLB face images from NIR images is rather studied, because VLB query image is effective for improving the accuracy of content-based image retrieval from the database consisting of VLB face images. For this task, colorization is not required. Chen *et al*. [5] proposed a method constructing VLB face image by block-wise linear regressions using similar patches with input image patches. Ma *et al*. [6] proposed a method estimating VLB image by representing input image as sparse linear combination of other NIR images. Colorization of NIR face images, however, is still important for crime investigation by human, we think.

3. Our colorization method

As we discussed in Section 1, our method consists of (1) position-dependent regression and (2) dimensionality reduction of independent variables by CCA.

Figure 1 Images of criminal behaviour detected by security camera at night.

†1 Wakayama University

For the position-dependent regression, we have to perform the image segmentation. The requirements of this segmentation are described below.

1. Pixels having similar monochrome intensity in the same face organ should belong to the same image segment.
2. Pixels having quite different intensities or pixels belonging to different organs should belong to different image segments.

These requirements are necessary for guaranteeing one-to-one mapping between intensity and color within each segment while keeping enough training samples for the regression.

This segmentation can be designed as two-stage processes as shown in Figure 2. In the first stage, a template named *part map* consisting of macro segments is constructed from color VLB face image set. Face images are initially divided into the macro segments in part map. In the second stage, the macro segments are further segmented into micro segments. Micro segmentation is performed depending on the intensity value of the NIR image so that the pixels in the same region have similar intensities. These two-stage segmentations reduce multiple correspondences between the intensity value of the NIR image and the color of the VLB image.

The independent variable in the above regression is the monochrome intensity value. This can be extended to a multi-dimensional vector by adding texture information around the pixel so as to estimate accurate color vector. However, too much extension may cause multi-collinearity problem, which may produce inaccurate results. For avoiding this problem, we also propose the dimensionality reduction by Canonical Correlation Analysis (CCA). Between the reduced variables, linear regressions are performed.

The details on the segmentation, regression with CCA, and colorization of unknown monochrome face image are described in the following subsections.

3.1 Face image segmentation

This section describes part map construction from face images and the segmentation of monochrome face images using part map. Hereafter, we assume almost aligned same-size VLB and NIR face image pairs are provided as training data.

3.1.1 Part map construction

Part map is the image consisting of macro segments representing the facial organs, such as, eyebrows, eyes, nose, mouth, cheeks, and so on. Part map can be designed manually, however, bottom-up construction from face images produces better map.

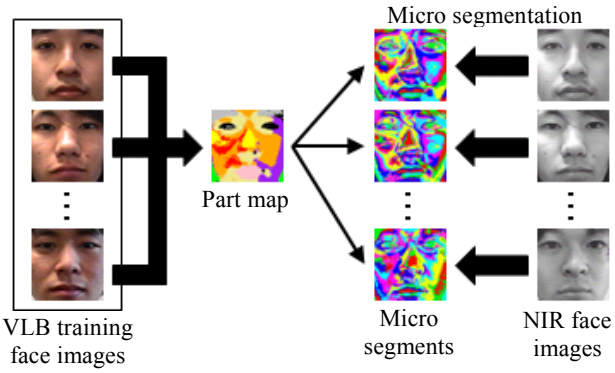


Figure 2 Two-stage segmentation.

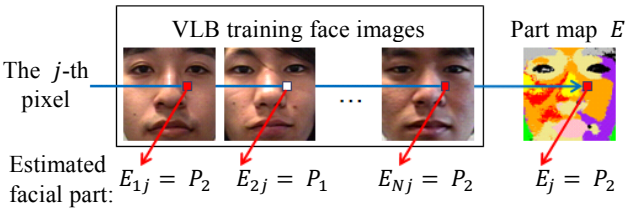


Figure 3 Construction of part map.

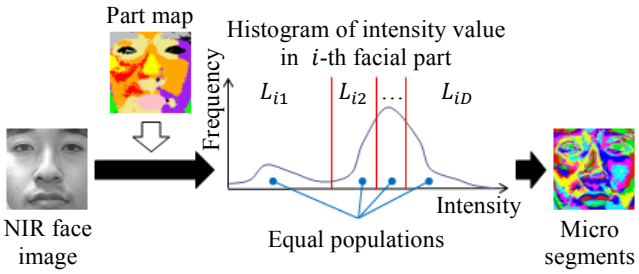


Figure 4 Micro segmentation on intensity value.

First, we provide color pixels belonging to each facial part from VLB training face images and the color distribution within the part is modelled as the following Gaussian:

$$g(\boldsymbol{x};\boldsymbol{\mu}_i,S_i)=\frac{1}{\left(\sqrt{2\pi}\right)^3\sqrt{|S_i|}}\exp\left(-\frac{(\boldsymbol{x}-\boldsymbol{\mu}_i)^T S_i^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_i)}{2}\right), \quad (1)$$

$$\boldsymbol{\mu}_i=\frac{1}{|P_i|}\sum_{\boldsymbol{x}\in P_i}\boldsymbol{x}, \quad (2)$$

$$S_i=\frac{1}{|P_i|}\sum_{\boldsymbol{x}\in P_i}(\boldsymbol{x}-\boldsymbol{\mu}_i)(\boldsymbol{x}-\boldsymbol{\mu}_i)^T, \quad (3)$$

where P_i $(i=1,2,\cdots,M)$ represents i-th facial part color set, $\boldsymbol{x}$ a color vector $(R \quad G \quad B)^T$, $\boldsymbol{\mu}_i$ and S_i the mean and the covariance matrix of P_i, and $|P_i|$ the number of colors in P_i.

For j-th pixel position, color $\boldsymbol{I}_{kj}$ $(k=1,2,\cdots,N)$ of the pixel is sampled from k-th training image. Then, compute the Gaussian of i-th part for the sampled color $\boldsymbol{I}_{kj}$. We denote the part E_{kj} providing the maximum Gaussian for the k-th image at j-th pixel as shown in Equation (4).

$$E_{kj}=\underset{i=1,2,\cdots,M}{\operatorname{argmax}}\left(g(\boldsymbol{I}_{kj};\boldsymbol{\mu}_i,S_i)\right). \quad (4)$$

The part E_j getting the highest frequency at j-th position is the value of part map at the pixel as shown in Figure 3. This can be expressed as

$$E_j=\underset{i=1,2,\cdots,M}{\operatorname{argmax}}\left(h(i,j)\right), \quad (5)$$

where $h(i,j)$ represents frequency of the i-th part at the j-th pixel. Part map E is constructed by solving E_j at all pixel locations.

3.1.2 Micro segmentation

Even if some pixels belong to a part, their intensities may have wide variations. For example, pixels belonging to nose part may be bright in specular reflection area, or dark in shadow area. Such signal-level variations may enlarge linear regression error. For relaxing this problem, we further segment the macro segments. This micro segmentation should depend on the monochrome image intensity. For guaranteeing enough population of pixels belonging to each micro segment, we

employ intensity histogram based segmentation as shown in Figure 4. Histogram intervals L_{il} $(l = 1, 2, \cdots, D)$ are obtained by dividing the histogram of i-th facial part into D parts having equal populations.

The micro segmentation is applied to all monochrome face images including training and test data.

3.2 Regression with dimensionality reduction

In our method, monochrome images are colorized by pixel-wise linear regressions. For this regression, one may think that richer independent variables produces better color estimation. In the discussion above, we only mentioned intensity value as an independent variable. However, this can be extended to multiple variables by adding texture information around the pixel. However, too many independent variables may cause multi-collinearity problem that may affect the regression accuracy. For avoiding this, dimensionality reduction is required. In our method, CCA [7][8][9][10] is employed for reducing the number of independent variables. The only drawback of this computation is that the bases obtained by CCA are non-orthogonal. Because of this, RGB values cannot simply be estimated by linear combination of output bases with dependent variable weight. For the simplification, obtained bases are converted into bi-orthogonal bases so that RGB values can be estimated by their linear combination.

3.2.1 CCA

This section describes CCA as the dimensionality reduction method based on maximization of squared canonical correlation between variables of data.

Let the vectors consisting of r and s variables $(r \leq s)$ be $\boldsymbol{X} = (x_1 \quad x_2 \quad \cdots \quad x_r)^T$ and $\boldsymbol{Y} = (y_1 \quad y_2 \quad \cdots \quad y_s)^T$ respectively. By enforcing $\mathrm{E}\{\boldsymbol{X}\} = \boldsymbol{0}$ and $\mathrm{E}\{\boldsymbol{Y}\} = \boldsymbol{0}$, covariance matrices are expressed as $\mathrm{V}\{\boldsymbol{X}\} = E\{\boldsymbol{X}\boldsymbol{X}^T\} \equiv S_{xx}$, $\mathrm{V}\{\boldsymbol{Y}\} = E\{\boldsymbol{Y}\boldsymbol{Y}^T\} \equiv S_{yy}$, and $\mathrm{cov}\{\boldsymbol{X}, \boldsymbol{Y}\} = \boldsymbol{E}\{\boldsymbol{X}\boldsymbol{Y}^T\} \equiv S_{xy}$.

z_x and z_y are obtained by projecting $\boldsymbol{X}$ and $\boldsymbol{Y}$ to bases $\boldsymbol{a}$ and $\boldsymbol{b}$ respectively.

$$z_x = \boldsymbol{a}^T\boldsymbol{X}, \tag{6}$$

$$z_y = \boldsymbol{b}^T\boldsymbol{Y}. \tag{7}$$

These variables are called canonical variable. CCA estimates the bases that maximize the correlation between canonical variables.

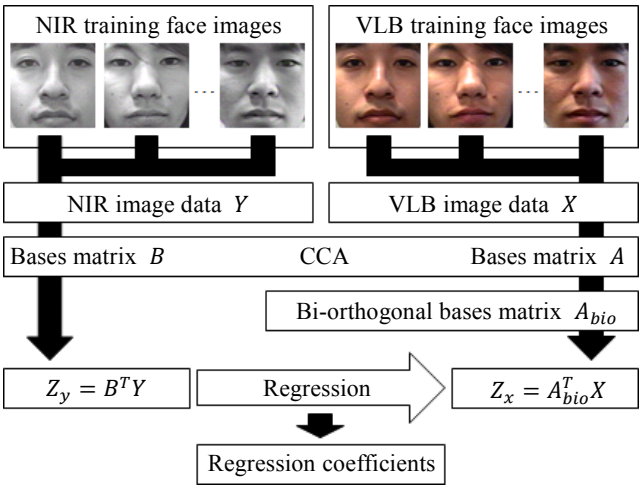


Figure 5 Training of face images.

The squared canonical correlation function ρ^2 is defined as

$$\rho^2 \equiv \frac{\mathrm{cov}^2\{z_x, z_y\}}{\mathrm{V}\{z_x\}\mathrm{V}\{z_y\}} = \frac{\left(\boldsymbol{a}^T S_{xy}\boldsymbol{b}\right)^2}{\boldsymbol{a}^T S_{xx}\boldsymbol{a}\boldsymbol{b}^T S_{yy}\boldsymbol{b}}. \tag{8}$$

The bases that maximize Equation (8) are obtained by solving the following simultaneous eigenproblem:

$$\begin{cases} S_{xx}^{-1}S_{xy}S_{yy}^{-1}S_{yx}\boldsymbol{a} = \rho^2\boldsymbol{a} \\ S_{yy}^{-1}S_{yx}S_{xx}^{-1}S_{xy}\boldsymbol{b} = \rho^2\boldsymbol{b} \end{cases}. \tag{9}$$

Bases vectors corresponding to maximum eigenvalue ρ_1^2 are $\boldsymbol{a}_1$ and $\boldsymbol{b}_1$, and they maximize the value of canonical correlation function ρ_1 between canonical variables $z_{x1} = \boldsymbol{a}_1^T\boldsymbol{X}$ and $z_{y1} = \boldsymbol{b}_1^T\boldsymbol{Y}$. The canonical variable obtained by using eigenvector (namely, basis vector) $\boldsymbol{a}_t$ and $\boldsymbol{b}_t$ corresponding to t-th eigenvalue is called t-th canonical variable.

r canonical variables are obtained because r eigenvalues are obtained. Therefore, s-dimensional data are reduced to r-dimensional data because r basis vectors are obtained.

3.2.2 Bi-orthogonal expansion

Bi-orthogonal expansion is the expansion of $\boldsymbol{w} = \sum_{m=1}^{q}(\boldsymbol{w}^T\boldsymbol{u}_m)\,\boldsymbol{v}_n$, where $\{\boldsymbol{u}_m\}_{m=1,2,\cdots,q}$ and $\{\boldsymbol{v}_n\}_{n=1,2,\cdots,q}$ represent bi-orthogonal bases and $\boldsymbol{w}$ a calculated vector. Bi-orthogonal bases satisfy the relation of $\boldsymbol{u}_m^T\boldsymbol{v}_n = \delta_{m,n}$, where $\delta_{m,n}$ represents the Kronecker delta. These bases can be expanded as well as orthogonal basis even if they are non-orthogonal bases.

Let A_{bio} and A be a matrix of bi-orthogonal bases and a matrix of bases with dependent variables obtained by CCA respectively. The relations of $A^TA_{bio} = I$ and $AA_{bio}^T = I$ are satisfied from the above relation, where I represents the identity matrix. Therefore, A_{bio} is the pseudo-inverse matrix of A^T.

3.2.3 Training between VLB and NIR face images

This section describes the training between NIR and VLB face images by multiple linear regressions as shown in Figure 5. In this report, let the dependent variable be 3-dimensional vector $\boldsymbol{x}$ (consisting of RGB value in VLB images). Let the independent variable be 8-dimensional vector $\boldsymbol{y}$ (consisting of the following texture information: $f(x,y)$, $\partial f/\partial x$, $\partial f/\partial y$, $\partial^2 f/\partial x^2$, $\partial^2 f/\partial y^2$, $\partial^2 f/\partial xy$, x, and y, where f represents the intensity in NIR images). The vectors are treated as matrices $X = (\boldsymbol{x}_1 \quad \boldsymbol{x}_2 \quad \cdots \quad \boldsymbol{x}_P)$ and $Y = (\boldsymbol{y}_1 \quad \boldsymbol{y}_2 \quad \cdots \quad \boldsymbol{y}_P)$ corresponding to VLB and NIR training face images respectively, where P represents the total number of pixels of training images. The mean values of each variable of these matrices are adjusted to 0 by subtracting the mean values of each variable from themselves.

First, the number of independent variables is reduced by using CCA. The bases obtained by CCA are treated as matrices $A = (\boldsymbol{a}_1 \quad \boldsymbol{a}_2 \quad \boldsymbol{a}_3)$ and $B = (\boldsymbol{b}_1 \quad \boldsymbol{b}_2 \quad \boldsymbol{b}_3)$. Therefore, the bi-orthogonal bases matrix corresponding to dependent variables is converted as $A_{bio} = A^{T^+}$. A_{bio} and B construct matrices $Z_x = A_{bio}^TX$ and $Z_y = B^TY$ respectively.

Then, the multiple linear regressions are performed from Z_y to Z_x within for each micro segment. For the same micro segment, independent variables are all variables of Z_y and the dependent variable is each variable of Z_x. In other words, the multiple linear regression is performed 3 times for one micro segment because Z_x consists of 3-dimentional vectors.

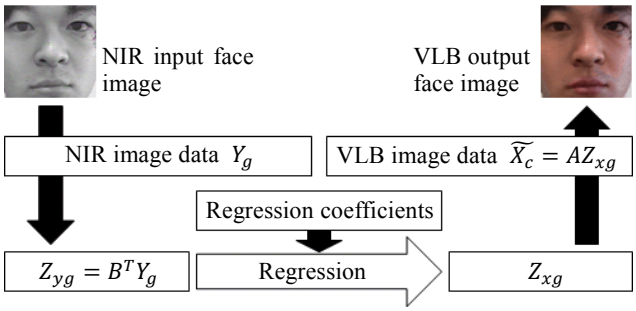


Figure 6 Colorization of an unknown NIR face image.

Finally, $M \times D$ sets of the regression coefficient are trained from all micro segments.

3.3 Colorization of an unknown NIR face image

This section describes the colorization of an unknown NIR input face image by regression for each micro segment by using trained regression coefficients as shown in Figure 6.

First, as well as training, let the NIR input face image be matrix Y_g consisting of 8-dimensional vectors. Projecting Y_g to bases matrix obtains matrix $Z_{yg} = B^T Y_g$.

Then, for each Z_{yg} within the same micro segment, regressions are performed by using regression coefficients trained in Section 3.2.3. Let Z_{xg} be a matrix of data obtained by all regressions.

Finally, the VLB output face image is constructed from Z_{xg} by using bi-orthogonal expansion. It can be considered that Z_{xg}satisfies the following relation:

$$Z_{xg} \approx A_{bio}^T X_c, \quad (10)$$

where X_c represents the matrix consisting of the 3-dimentional vectors, as well as training, of the true VLB face image. Therefore, the following equation holds:

$$\widetilde{X_c} = AZ_{xg}, \quad (11)$$

where $\widetilde{X_c}$ represents the estimated matrix for X_c.

The output matrix consisting of the vectors of RGB value is obtained by adding mean values, which have been subtracted for training, to each variable of $\widetilde{X_c}$.

4. Experiments

In order to confirm the effectiveness of our colorization method, we conducted two experiments.

Experiment 1. Comparison between colorization methods with and without two-stage segmentation using part map.

Experiment 2. Comparison between colorization methods with and without dimensionality reduction.

In the experiments, we used the dataset consists of 49 pairs of NIR and VLB face images taken by an off-the-shelf surveillance camera (TR-303WVP by Corona Dengyo Co., Ltd).

All experiments are conducted in one-out-of-many manner. That is, for colorizing a test monochrome image, part map construction and regression coefficient learning are performed on 48 images excluding the test image. Part map used in our method consists of 9 facial parts. Each part is divided into 4 micro segments. As a result, a face image is divided into 36 regions in total.

In Experiment 1, a colorization without using part-map-based macro segmentation was applied to the test data. In this case, the number of micro segments is set to 36 for keeping the fairness. We call this method *Naive*.

In Experiment 2, the original independent variables are 8-dimension consisting of intensity, first and second order differential variables, and coordinate values of x and y. The dependent variables are 3-dimension representing R, G, and B. In our method, by applying CCA, independent variables are reduced to 3-dimension. Without applying CCA, the colorization problem becomes a regression from 8 to 3 dimensional spaces. We implemented this regression as a reference.

Experiment 1 First, we compared between the colorized images constructed by our method and Naive. The colorization results are shown in Figure 7. The figure shows the input image, ground truth, output images constructed by our method and Naive, and micro segments constructed by our method and Naive. For the micro segments, the pseudo-colors denote the micro segments. From this figure, it seems that our method estimates more accurate colors than Naive. Especially, our method estimates vivid red lips compared with Naive, and the background region in Naive has skin color but our method estimates white background.

Figure 8 and 9 shows the color distributions of these four output images and true color distributions. The color distributions of the output images constructed by our method are slightly more similar to true color distributions than those distributions by Naive.

Table 1 shows the mean absolute error for H, S, and V color channels [11]. These values are obtained by pixel-wise comparison between ground truth and output images for 49 images. This table demonstrates that our method can colorize with smaller errors than Naive.

From this experiment, we can conclude that our method with two-stage segmentation estimates more accurate color images than the method with one-stage segmentation.

Experiment 2 In this report, CCA is used for dimensionality reduction as described in Section 3.2. Dimensionality reduction is required for avoiding multi-collinearity problem. Figure 10 shows (a) the result of regression from reduced 3 independent parameters to 3 dependent parameters, and (b) the result of regression from 8 independent variables to 3 dependent variables. This comparison strongly supports the necessity of the dimensionality reduction by CCA.

5. Conclusions

We presented a method that colorizes a monochrome NIR face image by regression with coefficients trained for each facial part. In this process, face images are divided by two-stage segmentation consisting of macro segmentation using part map and micro segmentation referring the intensity of the monochrome image. In our method, dimensionality reduction is also applied by using CCA for avoiding multi-collinearity problem. Bi-orthogonal bases are used for estimating the output color from dependent variables obtained by the regression. In the experiments, we confirmed the effectiveness of the part-map-based segmentation and the CCA based dimensionality reduction.

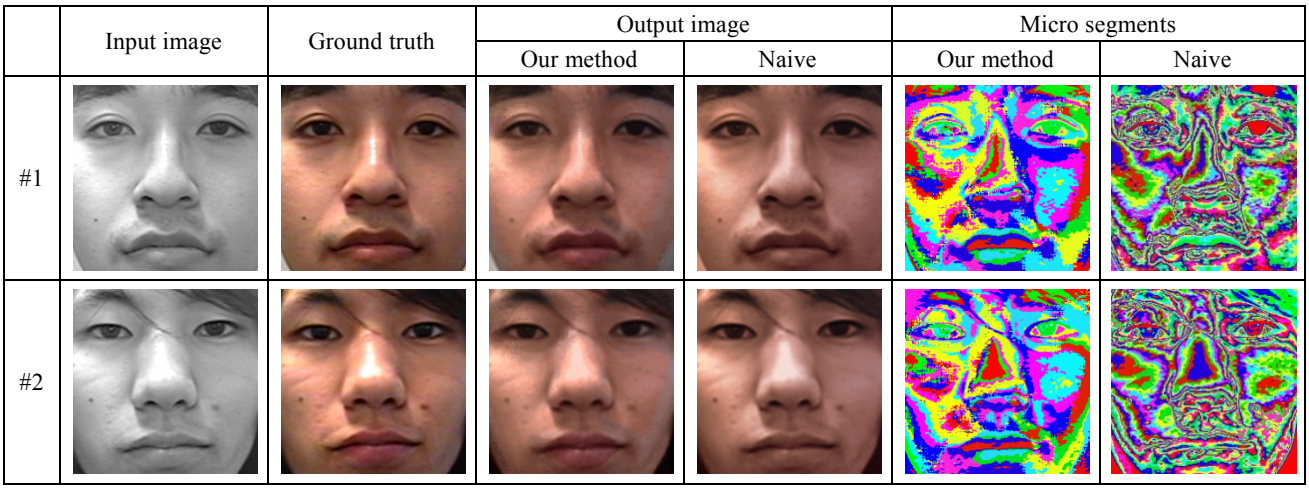
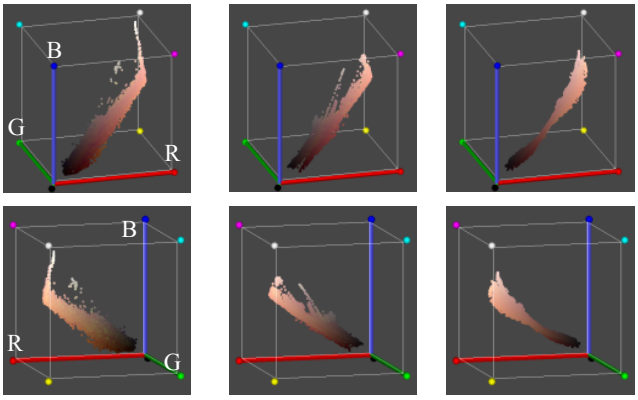
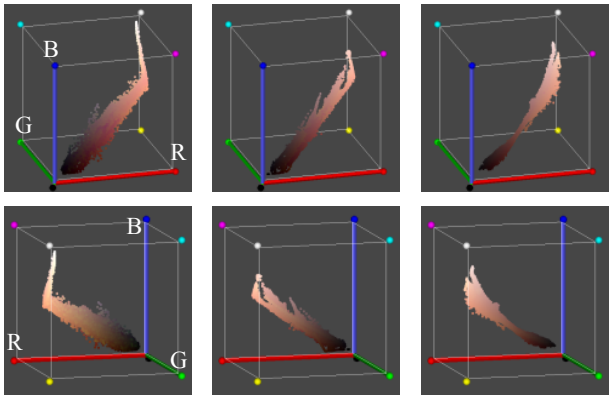


Figure 7 Results of the colorization.



True distribution Our method Naive

Figure 8 Color distribution of image #1.

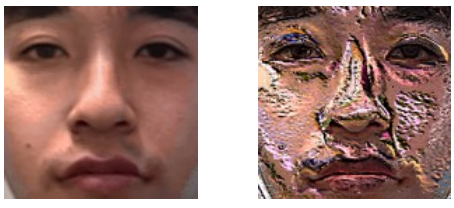


True distribution Our method Naive

Figure 9 Color distribution of image #2.

Table 1 The mean absolute error of the output images.

	Our method	Naive
H value	10.5601	11.2113
S value	0.0755	0.0846
V value	0.0578	0.0607



(a) (b)

Figure 10 The effect of dimensionality reduction: (a) Regression result from reduced 3D input vector to 3D output vector by CCA. (b) Regression result from 8D to 3D RGB vector.

Currently, our method has a limitation that output images depend not only on input monochrome image but also training images. This implies that we have to prepare the training images taken under similar illumination with input image. Also the face pose and skin colors in the training images have to be similar with the input images. This problem will be addressed in the future works.

Reference

[1] T. Horiuchi and S. Hirano, Colorization Algorithm for Grayscale Image by Propagating Seed Pixels, International Conference on Image Processing (ICIP 2003), pp.457-460, 2003.

[2] A. Levin, D. Lischinski and Y. Weiss, Colorization using Optimization, ACM Transactions on Graphics (TOG), Vol.23, 3, pp.689-694, 2004.

[3] R. Irony, D. Cohen-Or and D. Lischinski, Colorization by Example, Eurographics Symposium on Rendering, pp.201-210, 2005.

[4] X. Liu, L. Wan, Y. Qu, T.T. Wong, S. Lin, C.S. Leung and P.A. Heng, Intrinsic Colorization, ACM Transactions on Graphics (TOG), Vol.27, 5, No.152, 2008.

[5] J. Chen, D. Yi, J. Yang, G. Zhao, S.Z. Li and M. Pietikäinen, Learning Mappings for Face Synthesis from Near Infrared to Visual Light Images, Computer Vision and Pattern Recognition (CVPR 2009), pp.156-163, 2009.

[6] Y. Ma, W. Xie, X. Chen, M. Liu, Y. Guo, J. Meng, Z. Yuan and Q. Qin, An Analysis-by-Synthesis Method based on Sparse Representation for Heterogeneous Face Biometrics, International Conference on Multimedia Technology (ICMT 2011), pp.148-151, 2011.

[7] H. Hotelling, Relations between two sets of variates, Biometrika, 28, pp.321-377, 1936.

[8] David Weenink, Canonical correlation analysis, Proceedings of Institute of Phonetic Sciences of the University of Amsterdam 25, pp.81-99, 2003.

[9] M. Borga, T. Landelius and H. Kuntsson, A Unified Approach to PCA, PLS, MLR and CCA, Report LiTH-ISY-R-1992, SE-581 83 Linköping, 1997.

[10] M. Kawaguchi, An Introduction to Multivariate Analysis, Morikita Publishing Co., Ltd., pp.53-60, 2008.

[11] A.R. Smith, Color gamut transform pairs, Computer Graphics, vol.12, 3, pp.12-19, 1978.