Efficient and Fast Multi-View Face Detection Based on Feature Transformation

Dongyoon Han*, Jiwhan Kim*, Jeongwoo Ju*, Injae Lee**, Jihun Cha**, Junmo Kim*

*Department of EECS, Korea Advanced Institute of Science and Technology (KAIST),
373-1 Guseong-Dong, Yuseong-Gu, Daejeon 305-701, Republic of Korea

**Broadcasting & Telecommunications Media Research Laboratory, ETRI,
161 Gajeong-Dong, Yuseong-Gu, Deajeon, 305-350, Republic of Korea

calintz@kaist.ac.kr, jhkim89@kaist.ac.kr, veryju@kaist.ac.kr, ninja@etri.re.kr, jihun@etri.re.kr, junmo@ee.kaist.ac.kr

***Abstract*— The training time of Adaboost to obtain the strong classifier is usually time-consuming. Moreover, to deal with rotated faces, it is natural to need much more processing time for both training and execution stages. In this paper, we propose new efficient and fast multi-view face detection method based on Adaboost. From the robustness property of Harr-like feature, we first construct the strong classifier more effective to detect rotated face, and then we also propose new method that can reduce the training time. We call the method feature transformation method, which rotates and reflects entire weak classifiers of the strong classifier to construct new strong classifiers. Using our proposed feature transformation method, elapsed training time decrease significantly. We also test our face detectors on real-time HD images, and the results show the effectiveness of our proposed method.**

***Keywords*— Face Detection, Multi-view Face Detection, Haar-like Features, Feature Reflection and Rotation, Cascade Classifier**

I. INTRODUCTION

Face detection is a task of finding the location of face region from single image or video. Face detection can be used in biometrics, such as a pre-processing step of face recognition systems, robot vision applications, saliency detection, skin detection, surveillance, etc. Most face detectors are based on sub-window searching algorithm to find face regions. Several methods have been tried to determine whether a face is placed inside a sub-window or not. For example, neural-network based [3, 6] and support vector machine [5] had been used and got effective results. Viola et al. [8] proposed Adaboost [2] based face detection system with integral image based Haar-like features, which can be processed with high accuracy in real time.

However, it is hard to detect some rotated face regions by using the original Haar-like feature and its variants [7, 8] based face detectors. To overcome this limitation, there are several approaches to detect rotated faces. Some methods [1, 4, 9] used multiple face detectors which can only detect rotated faces with fixed angles with low execution time. The other methods [3, 6] based on neural-network often shows fine performance with robustness to rotation, but they commonly have relatively complex framework and most of them require long training and execution time for real-time implementation. Especially in [4], entire input image is rotated twice to detect rotated faces, which needs to generate integral images according to the number of rotation, so the process is obviously time-consuming and redundant.

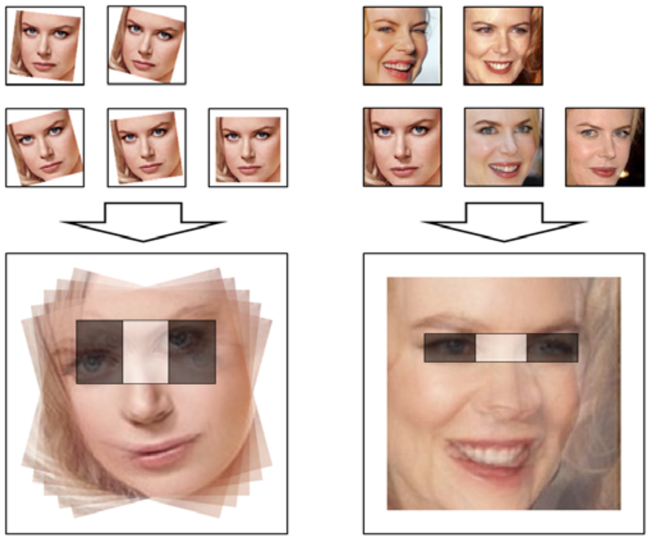

Figure 1. A toy example depicting the robustness of Haar-like features under small in-plane and out-of-plane rotations. It is shown that each third type Haar-like feature both discover the eye region well through the in-plane and out-of-plane rotated faces well, respectively.

The rest of the paper is organized as follows. In Section II, we describe the process of detecting a face region by using Adaboost. We also introduce the feature reflection and rotation method for rotation invariant face detection system. Experimental results are presented in Section III. In Section IV, we conclude our work with further discussion.

II. PROPOSED MULTI-VIEW FACE DETECTION

A. *Robustness of Haar-like features*

Widely used Haar-like features [8] are shown in Figure 2, we call the features first, second, third, fourth, and fifth type feature from the left. One of great advantages of using Haar-like features chosen for the features of Adaboost is the robustness of features. Because each output value of Haar-like features are calculated by the average pixel intensities inside rectangular areas by the integral images [8], slight changes of faces in training face set are neglible to learn informative Haar-like features. In other words, Haar-like features conduct like rectangular average filter, output of which is not readily varying by small changes of input.

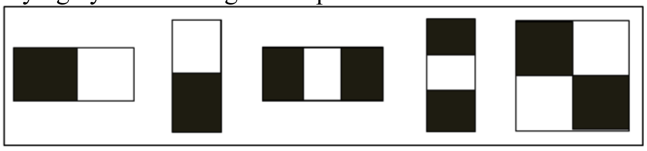

Figure 2. Five kind of widely used Haar-like features. From the left, first, second, third, fourth, and fifth features are presented.

Thus, Haar-like feature based Adaboost can effectively learn to construct the best informative weak classifiers from a great number of diverse faces and the convergence of Adaboost is also preserved by the robustness. For instance, positions and sizes of eyes of each diverse upright face are different, but a third type Haar-like feature can cover the region under small variances of faces. Figure 1 is a toy example to show the robustness of a Haar-like feature. We can notice that some slight changes of the in-plane angle and out-of-plane angle cannot affect the training process too much.

The robustness of Haar-like features can be utilized to deal with rotated faces in the training stage. From a given original training set consisting of several faces, we enlarge the training set by rotating them to construct an expanded training set. Then, in-plane rotated faces can contribute to detect similar view faces in the test stage. Out-of-plane rotated faces cannot be easily produced by rigid transformations such as rotation, so it is necessary to find the faces captured in different viewpoints corresponding to the out-of-plane rotation to construct the original training set. Note that this process is the in-plane process, but if the training set contains several angles of out-of-plane rotated faces inherently, expanded faces also contribute to detect out-of-plane rotate faces in the test stage.

However, the rotations degree of additional faces should be constrained inside a proper range. Because, high degrees of rotation of faces cause large intra-variance of the positive samples, so that Adaboost cannot converge well to classify the positive and negative samples. Besides, the convergence rate of Adaboost drops proportionally to the rotation angle and it diverges in the long run.

We determined that the proper range of in-plane rotation is from -15 degrees to +15 degrees and out-of-plane rotation is from -15 degrees to +15 degrees of nearly upright faces. This is set by the cross-validation of the convergence rate of Adaboost in training stage and the discriminative power of the learned strong classifier. Under the setting we mentioned above and the expanded nearly upright (0 degree) face training set, we can obtain the strong classifier, which can detect from -15 degrees to +15 degrees in-plane and out–plane rotated faces. If we learn a strong classifier from the expanded training set, original training set of which consists of about 30 degree in-pane rotated faces, we can detect from 15 degrees to 45 degree in-plane rotated faces.

B. *Feature transformation*

We propose the feature transformation method to obtain the strong classifier from the existing strong classifier for detecting rotated faces. Following the method in previous section, a $w \times w$ size face detector with strong classifier learned from expanded training set can detect particular range of rotated faces, but not entire angle range. Hence, to cover more range, we need more classifiers. A simple way to increase the number of classifiers that can cover entire angle range is to train n classifiers from each corresponding rotated training set. However, this approach is time-consuming, since the training time increases proportionally to the number of classifiers.

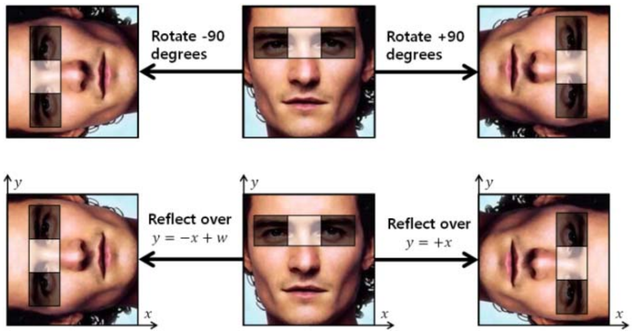

Figure 3. A toy example illustrating our proposed feature transformation method. The third type Haar-like feature on upright face is reflected or rotated to detect +90 and -90 degrees faces.

Let C_0 denotes the strong classifiers learned from the nearly upright face training set, we can produce two strong classifiers C_{+90} and C_{-90}, which can detect +90 degrees and -90 degrees in-plane rotated faces, respectively. This is done by feature transformation: feature reflection and rotation. Haar-like features of the upright face can be reused on the ±90 degrees in-plane rotated faces with simple coordinate changes and type changes. As shown in Figure 3, if we reflect the coordinate and the type of the third type Haar-like feature on the upright face, we can detect ±90 degrees in-plane rotated faces without learning additional strong classifiers. We also notice that the difference between the reflection and the rotation is almost negligible, because the face is symmetric, so most of the learned weak classifiers are placed symmetrically in general. To calculate the value of new weak classifiers using integral image, sometimes black and white regions are exchanged and the sign of the values are also reversed. We denote these ‘feature region preservation’ and ‘sign of feature region values’, respectively. The detailed method is described in Table 1.

We tested both rotated and reflected strong classifiers, but the performance is also almost equivalent. The strong classifiers C_0, C_{+90} and C_{-90} can cover from -15 degrees to 15 degrees, from 75 degrees to 105 degrees and from -75 degrees to – 105 degrees, respectively.

TABLE 1. THE FEATURE ROTATION METHOD FOR CREATING NEW +90 AND -90 STRONG CLASSIFIERS.

Method 1: Feature rotation method		
C_{+90}	Feature coordinate change	$y \leftarrow w - x, x \leftarrow y$
	Feature type change	First type ↔ Second type Third type ↔ Fourth type Fifth type ↔ Fifth type
	Sign of feature region values	Reversed
	Feature region preservation	For first, second, third, fourth type features, region is preserved, but for fifth type feature, black and white region should be exchanged.
C_{-90}	Feature coordinate change	$y \leftarrow x, x \leftarrow w - y$
	Feature type change	First type ↔ Second type Third type ↔ Fourth type Fifth type ↔ Fifth type
	Sign of feature region values	Reversed
	Feature region preservation	For first, second, third, fourth type features, region is preserved, but for fifth type feature, black and white region should be exchanged.

We now extend our proposed method to cover other angle ranges (i.e., from -15 degrees to -75 degrees and from +15 degrees to +75 degrees). On the previous feature transform method, we start from one base strong classifier for upright faces to produce two other classifiers for +90 and -90 degrees rotated faces. We now start from another base strong classifier for +30 degrees rotated faces. The strong classifier is learned from +30 degrees rotated and cropped faces of original training faces. As shown in Figure 4 if we reflect the coordinate and the type of the Haar-like features on the +30 degrees rotated face, we can detect -30 and ±60 degrees in-plane rotated faces without learning additional strong classifiers. Note that in this case, we need not only rotations, but reflections of features to detect reverse degrees of given base strong classifier. The detailed method is described in Table 2.

TABLE 2. THE FEATURE ROTATION METHOD FOR CREATING NEW +90 AND -90 STRONG CLASSIFIERS.

Method 2: Feature transform method		
C_{-60}	Transform method	Rotation
	Feature coordinate change	$y \leftarrow x, x \leftarrow w - y$
	Feature type change	First type ↔ Second type Third type ↔ Fourth type Fifth type ↔ Fifth type
	Sign of feature region values	Reversed
	Feature region preservation	For first, second, third, fourth type features, region is preserved, but for fifth type feature, black and white region should be exchanged.
C_{-30}	Transform method	Reflection
	Feature coordinate change	$y \leftarrow y, x \leftarrow w - x$
	Feature type change	Preserved
	Sign of feature region values	For the values of first and fifth features should be reversed.
	Feature region preservation	For second, third, fourth type features, region is preserved, but for first and fifth type features, black and white region should be exchanged.
C_{+60}	Transform method	Reflection
	Feature coordinate change	$y \leftarrow x, x \leftarrow y$
	Feature type change	First type ↔ Second type Third type ↔ Fourth type Fifth type ↔ Fifth type
	Sign of feature region values	Preserved
	Feature region preservation	Preserved

The strongest advantages of the feature transform method are as follows. First, all rotation detections are conducted on single integral image. In other words, given upright Haar-like feature based face detector, an integral image must be generated to calculate the values of Haar-like feature, we use the integral image for the purpose of calculating feature values of rotated faces. Since weak classifiers are reflected or rotated, we don't need to generate any additional integral images for rotated images. In addition, the training time can be dramatically reduced in the training stage of Adaboost. Learning large numbers of weak classifiers is usually time-consuming, and if we need further strong classifiers for particular degrees to detect multi-view faces, the number of overall weak classifiers learned in the training stage is too excessive and redundant. Using our proposed feature transform method, we just need two strong classifiers, upright and +30 degrees face strong classifier and then we transform the strong classifiers to generate new strong classifiers very fast. Note that out-of-plane rotation is covered in the training stage by several training out-of-plane rotated faces that have different views as we mentioned above. Therefore, all classifiers can cover entire in-plane and specified out-of-plane angle range.

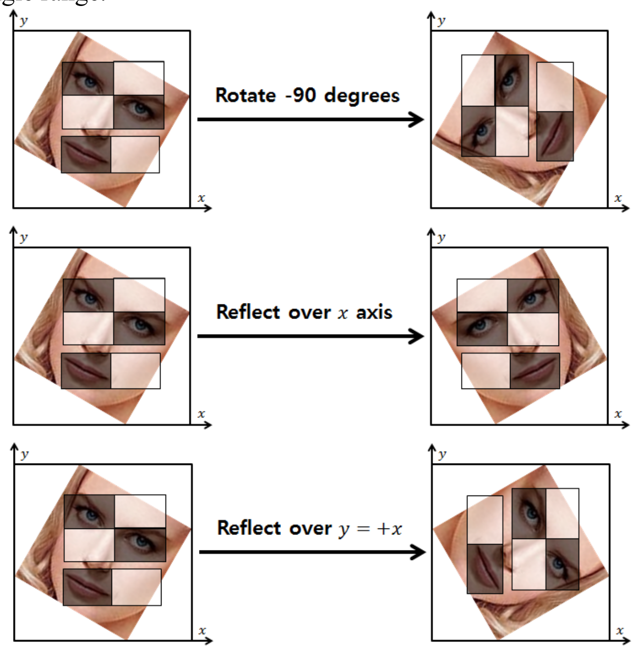

Figure 4. A toy example illustrating extended feature transformation method. The second and fifth type Haar-like features on +30 degrees face are reflected or rotated to detect -30, -60 and +60 degrees faces.

C. *Overall cascade structure*

Each of strong classifiers produced by our proposed feature transform method has a cascade structure themselves consisting of weak classifiers. Therefore, entire weak classifiers can be combined together to construct single unified strong classifier. Let n_{C_i} denotes the number of classifier stages, which comprise the strong classifier C_i. The overall unified strong classifier has $\sum_{i=1}^{n} n_{C_i}$ number of classifier stages as shown in Figure 5. On each stage, an input sub-window is determined whether face or not. We set each strong classifier with C_0, C_{+30},C_{-30} , C_{+60} , C_{-30} in sequence.

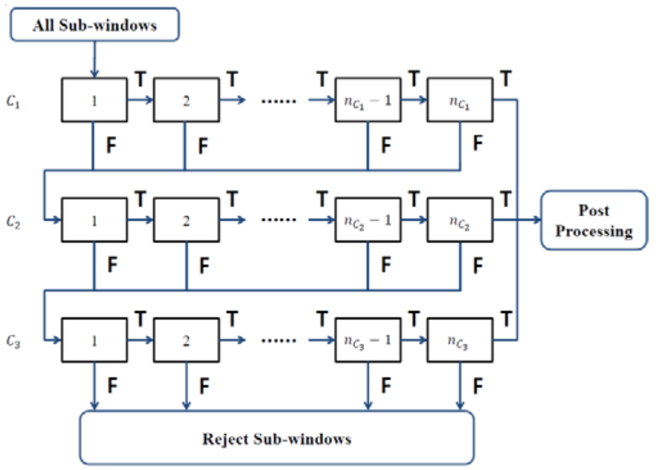

Figure 5. Overall cascade structure of combined strong classifers with three base strong classfiers.

III. Experimental Results

We tested our proposed methods with Haar-like Local Binary Pattern (HLBP) features [7], which are more robust to the illumination change than conventional Haar-like features. We train about 20,000 48×48 size face images and about 180,000 non-face images. Test and training environment is I5-760 CPU and 12GB RAM PC. We use two base strong classifiers: upright face strong classifier and +30 degree face strong classifier, and then we apply our proposed method to generate other strong classifiers. We first compare the training time that elapsed for generating whole strong classifiers with the training time that elapsed with our proposed method, and then we show the results on the real-time HD images.

A. *Training time comparison*

We compared our proposed feature transform method with the brute-force method that learns entire strong classifiers. Total elapsed time is measured under the same number of 2,800 features. It is shown that our proposed method significantly reduces the training time. After learning the base strong classifiers, our proposed method does not need additional time to learn new strong classifiers, so the feature learning time is much smaller than the whole classifiers learning method.

TABLE 3. Training time comparison

Method	# of features	Total elapsed training time	# of features trained per an hour
Learning entire classifiers	2,800	About two weeks	8.3
Feature transform method	2,800	46 hours	60.9

B. *Results on the real-time HD cam images*

We apply our strong classifiers to the real-time HD cam images. Figure 6 shows some test results. We can notice that the strong classifiers effectively detect in-plane and out-of-plane rotation faces.

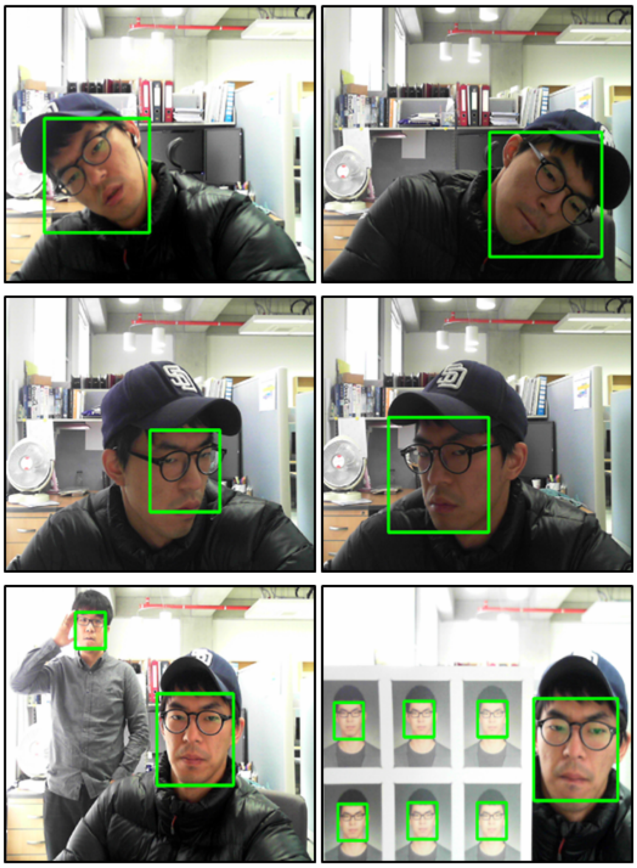

Figure 6. Example detection results on the real-time HD cam images.

IV. Conclusions

In this paper, we have proposed a new efficient and fast multi-view face detection method with the following new aspects. We have introduced a new method that can construct the effective strong classifiers to detect rotated faces from the robustness property of Harr-like features. Furthermore, we have also proposed a new method: 'feature transform method' that can reduce the training time. The experimental results demonstrate the effectiveness of our proposed method. For future work, we will generalize our feature transform method to other feature based face detectors such as MCT or LBP based face detector.

ACKNOWLEDGMENT

This research was funded by the MSIP (Ministry of Science, ICT & Future Planning), Korea in the ICT R&D Program 2013 (KCA-2013-(13-912-03-001)).

REFERENCES

[1] S. Du, N. Zheng, Q. You, Y. Wu, M. Yuan, and J. Wu, "Rotated haar-like features for face detection with in-plane rotation," Interactive Technologies and Sociotechnical Systems. Springer Berlin Heidelberg, pp. 128-137, 2006.

[2] Y. Freund, "Boosting a Weak Learning Algorithm by Majority Algorithm," Machine Leaning, vol. 43, no. 3, pp. 293-318, 2001.

[3] C. Huang, H. Ai, Y. Li and S. Lao, "High-Performance Rotation Invariant Multiview Face Detection," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 29, no. 4, April. 2007.

[4] B. Li, Y. Aiming, and J. Yang. "Rotated Face Detection Using AdaBoost," Proc. IEEE Information Engineering and Computer Science (ICIECS), 2010.

[5] E. Osuna, R. Freund, and F. Girosi, "Training Support Vector Machines: An Application to Face Detection," Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 130-136, 1997.

[6] H. Rowley, S. Baluja, and T. Kanade, "Rotation Invariant Neural Network-Based Face Detection," Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 38-44, 1998.

[7] A. Roy, S. Marcel, "Haar Local Binary Pattern Feature for Fast Illumination Invariant Face Detection". Proc. IEEE Conf. British Machine Vision Conference, pp. 1-12, 2009.

[8] P. Viola and M. Jones, "Rapid Object Detection Using a Boosted Cascade of Simple Features," Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2001.

[9] B. Wu, H. Ai, C. Huang, and S. Lao, "Fast Rotation Invariant Multi-View Face Detection Based on Real AdaBoost," Proc. Sixth Int'l Conf. Automatic Face and Gesture Recognition, pp. 79-84, 2004.

Dongyoon Han received the B.S. degree and the M.S. degree in Electrical Engineering and Computer Science from Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea, in 2011 and 2013, respectively, and is currently pursuing the Ph.D. degree in Electrical Engineering and Computer Science from Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea. His research interests are in machine learning, and computer vision.

Jiwhan Kim received his B.S. degree in Electronic Engineering in 2011 from Tsinghua University, Beijing, China and MS degree in Electrical Engineering from Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea, in 2013. Currently, he participates in Ph.D course in Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea. His research interests include saliency detection, face analysis, and pattern recognition.

Jeongwoo Ju received his B.S. degree in Aerospace engineering from Chungnam National University Daejeon, Korea in 2010 and M.S. degree in Division of Future Vehicle from Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea, in 2013. Currently, he participates in Ph.D course in Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea. His research interests include Deep Learning and face recognition.

Injae Lee received her B.S. and M.S. degrees in Electrical and Computer Engineering from Sungkyunkwan University, Suwon, Korea, in 1999 and 2001 respectively. Since 2001, she has been a senior member of research staff in Broadcasting &Telecommunications Media Research Division, Electronics and Telecommunications Research Institute (ETRI), S. Korea. Currently, he participates in the development of interactive rich media service technologies. Her research interests include interactive multimedia broadcasting system, interactive rich media system, 3D image processing and computer graphics.

Jihun Cha received his B.S. degree in Computer Science in 1993 from Myongji University, Yongin, Korea and M.S. and Ph.D. in Computer Science from Florida Institute of Technology, Melbourne, USA, in 1996 and 2002 respectively. He joined Electronics and Telecommunications Research Institute (ETRI) of Korea in 2003. He has been participated in a project on the development of Interactive Scalable Multimedia Streaming (ISMuS) platform and currently, participates in the development of T-DMB interactive multimedia service technologies. His research interests include multimedia streaming, interactive broadcasting system, and feature extraction in motion picture.

Junmo Kim received the B.S. degree from Seoul National University, Seoul, Korea, in 1998, and the M.S. and Ph.D. degrees from the Massachusetts Institute of Technology (MIT), Cambridge, in 2000 and 2005, respectively. From 2005 to 2009, he was with the Samsung Advanced Institute of Technology (SAIT), Korea, as a Research Staff Member. He joined the faculty of Korea Advanced Institute of Science and Technology (KAIST) in 2009, where he is currently an Assistant Professor of electrical engineering. His research interests are in image processing, computer vision, statistical signal processing, and information theory.