Wi-Fi Imaging Based Segmentation and Recognition of Continuous Activity

Yang Zi, Wei Xi(✉), Li Zhu, Fan Yu, Kun Zhao, and Zhi Wang

Xi'an Jiaotong University, Xi'an, China
ziyang783282949007@gmail.com, weixi.cs@gmail.com, zhuli@gmail.com, fanfanyyy1997@gmail.com, pandazhao1982@gmail.com, zhiwang.xjtu@gmail.com

Abstract. Automatic segmentation and action recognition have been a long-standing problem in sensorless sensing. In this paper, we propose CHAR, a continuous human activity recognition system to solve these problems in a different way. We've noticed that these challenges have been solved in image processing field, so CHAR could effectively perform action segmentation and recognition after WiFi imaging. The key idea behind Wi-Fi imaging is that different body part reflects transmitted signal, the receiver receives the combination of them, and then we separate the received signals from different directions and get the signal intensity in each direction to get the heat map showing the shape of the object. The imaging sequence contains multiple pictures records a continuous action at different time, and we can easily separate and recognize the action based on IC^2(image classification), a classification framework we proposed. We implement CHAR using commodity WiFi devices to evaluate its performance under different environment. The results show that the imaging result is better than prior works, facilitating CHAR to achieving an average recognition accuracy, i.e., >95%.

Keywords: Activity recognition · CSI · Wi-Fi imaging

1 Introduction

Human activity recognition is an importance technic in current applications, such as the human-computer interaction, somatic game, and health-care. Recent solutions fall into three categories: camera-based [1], sensor-based [2,3] and wireless-based [4,5] approaches.

Camera based approaches are able to guarantee high resolution for activity recognition. However, those approaches have fundamental limitations, including the line-of-sight detection, good illumination, and potential privacy leakage. On the other hand, sensor-based approaches usually require targets to carry on sensors, which is inconvenient in daily usage. Different from above solutions, leveraging wireless signals to achieve device-free activity recognition becomes promising, such as WiSee [4], E-eyes [6], and WiHear [7].

X. Wang et al. (Eds.): CollaborateCom 2019, LNICST 292, pp. 623–641, 2019.
https://doi.org/10.1007/978-3-030-30146-0_42

Those approaches are based on the observation that different human activities introduce different multi-path distortions in wireless signals, which can be used as the fingerprints of those activities. Nevertheless, there are still two drawbacks on the wireless signal based approaches. First, they usually can only distinguish activities in coarse granularity, e.g., [8,9]. Moreover, they often request specific facilities (e.g.USRP [4], GPS clock [10] or RFID [11,12]) to eliminate the impacts of ambient noises.

Recent advance in the research of WiFi networks proposes to utilize the Channel State Information(CSI) to realize fine-grained fingerprinting for activity recognition. CSI is sensitive to channel variances and position changes, which makes itself possible to capture the change as the experimenter performs action. However, CSI fingerprint based device-free activity recognition remains challenging. First, to perform continuous activity segmentation using CSI is extremely difficult. Second, CSI reflects the change of the channel, but its changes are difficult to match the corresponding specific movements. So when the receiver receives a continuous signal which contains two or more actions, it is difficult to distinguish them.

Another challenge for fingerprint based activity recognition is the device incompatibility. Due to the imperfect manufacturing process, different devices exhibit different signal gains. The variant gains make different CSI values once we change transmitting or receiving devices to detect the same activity. Hence, if some devices are changed, it is necessary to retrain the model for updating the fingerprint database.

The third challenge is to eliminate random disturbance caused by environmental noises and electromagnetic interferences. These two negative factors may result in unpredictable errors. Since the errors do not follow specific distributions, it is hard to eliminate or zero them by repeating trainings. In other words, even if a user performs a standard action identical to the one operated in the training phase, the CSI may still have a large difference from the fingerprint in the database.

In this paper, we propose a novel approach to solve the 3 aforementioned challenges. We've noticed that these challenges have been solved in image processing field, so is it applicable in our research? The answer is Yes and irrelevant to the existing fingerprint approaches. Instead, we propose a novel approach to perform Wi-Fi imaging first on which highly precise human activity recognition is implemented afterwards. The difficulty is how to perform WiFi imaging. Our basic idea is similar to optical imaging systems where images are typically formed by measuring the incoming signal intensities from each azimuth and elevation angle. Therefore, in our perception region, if we can get the signal strength from every direction, then we will get a heat map shows the shape of the object. After we obtain the imaging sequence using the phase shift across antennas, we can easily split continuous action imaging sequence. In order to better classify the heat map, we propose a new classification method called IC^2. The final classification result can be obtained from the IC^2.

Our contributions are summarized as follows:

1. CHAR proposes a novel approach to perform imaging using Wi-Fi signals and achieves preferable effect.

2. CHAR solves the problem of continuous motion segmentation by Wi-Fi imaging.
3. CHAR builds a bridge between wireless and pictures. Our extensive experiments show that CHAR is highly accurate in action recognition and insensitive to the diversity of individual users.

2 Related Work

In the literature, researches related to human activity recognition can be divided into two categories: Received Signal Strength Indicator (RSSI) based and Channel State Information (CSI) based approaches. In classification region, SVM (support vector machine) and CNN (convolutional neural networks) have been widely used and proved to have good performance.

2.1 RSSI Based

RSSI is sensitive to ambient movements, allowing it to produce a set of patterns for identifying locations [13] and human activities [14–16]. The work proposed in [14] employs RSSI measurements to obtain images of moving objects. The authors in [17] use kernel distance-based radio tomographic to locate a moving or stationary person. The authors in [18] design an RSSI based recognition system over mobile phones, identifying 7 different gestures. RSSI based recognition systems usually fail to recognize delicate motions because RSSI is too coarse to perform fine-grained detections [18].

2.2 CSI Based

CSI is also susceptible to human activities, such as walking, falling, presence and movements of part of human body. Because of its fine-grained WiFi signature, CSI is capable to support highly accurate activity recognition. Utilizing CSI, WiHear [7] detects lip and mouth movements. E-eyes [6] recognizes a set of human activities by leveraging CSI values as fingerprints. FCC [19] achieves device-free crowd counting using CSI. WiFall [8] detects people falls using CSI. The authors in [20] propose a stationary presence and mobile user detection scheme. CARM [9] utilizes the amplitude of CSI to recognize activities. ARM [10] uses both amplitude and phase of CSI to achieve gesture recognition.

2.3 Classification

Classification is one of the most active research and application areas of machine learning. The literature is vast and growing [21]. Traditional classification approaches, such as SVM (Support Vector Machine), DT (Decision Tree), have been widely applied for classification tasks, and exhibit great performance [21]. With the advent of convolutional neural networks (CNN), many researchers use it for classification problems. The work proposed in [22] employs CNN to Sentence

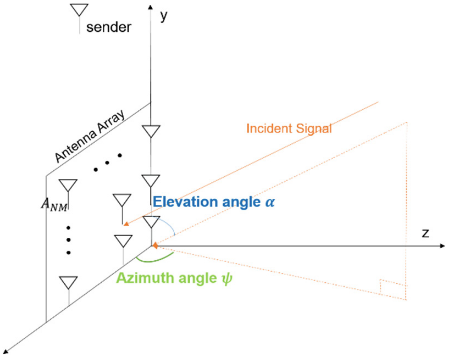


Fig. 1. Imaging system

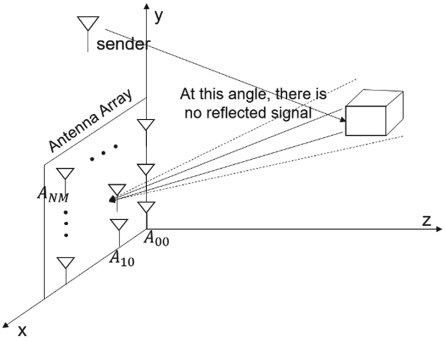


Fig. 2. Overview of imaging approach

Classification. The authors in [23] design a Classifier to Image Classification. In recent years, researchers have tried to achieve better classification performance by increasing the depth of CNN. The VGG [24] uses a 19-layer neural network and the Resnet [25] uses more than one hundred layers of network structure. Apart from the factor of depth, researchers have proposed other different aspect of architecture design, such as STN [26] and CBAM [27]. These modules can be inserted into existing convolutional architecture, and achieve better performance (Fig. 2).

3 Preliminary

In IEEE 802.11n standard, wireless communication uses OFDM modulated signals, which are transmitted over multiple orthogonal subcarriers, and each subcarrier have different frequencies [28]. For one subcarrier of frequency, the transmitted model in frequency domain can be expressed as:

$$Y(f) = H(f) \times X(f) + N(f). \tag{1}$$

Where $X(f)$ is the signal transmitted on subscarrier f, $Y(f)$ is the received signal, $N(f)$ is the additive white Gaussian noise vector and is the channel estimated result. If we have P subcarriers, we can get channel matrix $H = H(f)_{f=1...p}$ which is called the Channel State Information (CSI). CSI reflects the environment influences to the signal includes amplitude attenuation and phase shift. That is to say, the CSI phase measures the phase shift of the WiFi link between the transmitter and the receiver. What's more, the CSI can be easily obtained by COTS Intel 5300 NIC [29].

4 Design

In this section, we describe the processing flow of CHAR and address the associated challenges. CHAR includes the three main stages: WiFi imaging using CSI information received by commercial NICs, continuous action segmentation of image sequences, action recognition using IC^2.

4.1 CHAR's Imaging Algorithm

In this paper, we propose a novel approach to perform imaging using Wi-Fi signals. CHAR's approach is similar to optical imaging systems where images are typically formed by measuring the incoming signal intensities from each azimuth and elevation angle [30]. That is the transmitted signal can effectively "light up" reflective objects and the receiver uses the reflections to image the objects. Hence there is no need for distance computation and it can be implemented on commercial Wi-Fi APs. However this is not easy to accomplish in practice. because the receiver receives a linear of combination of reflections from multiple regions representing different body parts on each of its antennas. In an optical system, a lens is used to physically separate the received signals from different directions. CHAR, in contrast, uses multiple antennas and phase differences analysis to separate signals. In the rest of this section, we first recommend our image system which includes a two-dimensional antenna array as receiver and a directional antenna as transmitter, and then describe our image algorithm.

System Construction. CHAR is a system that captures human figure at first and then conduct activity recognition using these figures. The whole process includes transmitting Wi-Fi OFDM signals, receiving the reflections from different body parts, and processing these reflections to capture the human figure. CHAR's prototype consists of a directional antenna as transmitter and a two-dimensional antenna array as receiver as shown in Fig. 1. The antenna arrays along the x-y plane, and the antenna is located at the origin. There are a total of N and M antennas along the x-axis and y-axis respectively, of which the distance between two adjacent is d. To describe the direction of a reflected signal which can be received by the antenna array, two parameters are necessary. First, the angle between the signal and the X axis called azimuth angle. Second, the angle between the signal and the x-z plane called elevation angle.

CHAR's Imaging Algorithm. CHAR performs imaging using multiple antennas as Wi-Fi receiver which receives a linear combination of the multiple reflections from different directions, in other words, from different body parts (i.e., azimuth and elevation angles). Therefore, our key idea is to separate the received signals from different directions and get the signal intensity in each direction.

Consider a reflection signal $S(\Psi_k, \alpha_k)$ from the kth propagation path which represent a signal coming from a part of the body, arrives at the receiver from the azimuthal angle and the elevation angle α_k. The complex attenuation at the antenna in the origin of the signal after traveling along kth propagation path is denoted by γ_k. The attenuation at the second antenna in the array is the same except for an additional phase shift accumulated due to additional distance traveled by the signal which depends on d, Ψ_k and α_k.

Take two antenna A_{00} and A_{nm} of our antenna array as an example, we compute the phase shift between them. From basic physics, a distance difference Δd will introduce a phase shift $e^{-j\frac{2\pi \Delta d}{\lambda}}$, where λ is the signal wavelength.

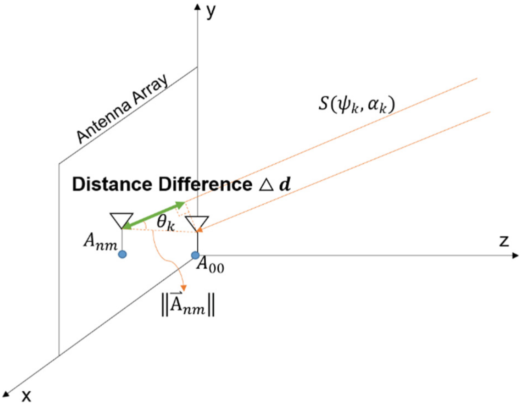


Fig. 3. Calculate the phase shift between antenna A_{00} and A_{nm}

Thus, as shown in Fig. 3, for signal $S(\psi_k, \alpha_k)$, the phase shift between antenna A_{00} and A_{nm} is given by:

$$\Phi_{n,m}(\psi_k, \alpha_k) = \gamma_k e^{-j\frac{2\pi \Delta d_{n,m}(\psi_k, \alpha_k)}{\lambda}} \tag{2}$$

where $\Delta d_{n,m}(\psi_k, \alpha_k)$ is the distance difference traveled by the signal between A_{00} and A_{nm}, as shown in the Fig. 3. According to trigonometric identities, we can derive the following equations:

$$\Delta d_{n,m}(\psi_k, \alpha_k) = \| \boldsymbol{A}_{nm} \| cos(\theta_k) \tag{3}$$

$$cos(\theta_k) = \frac{\boldsymbol{S}(\psi_k, \alpha_k) \cdot \boldsymbol{A}_{nm}}{\| \boldsymbol{S}(\psi_k, \alpha_k) \| \| \boldsymbol{A}_{nm} \|} \tag{4}$$

Where θ_k is the angle between the signal and the x-y plane, $\boldsymbol{A}_{nm}$ is the vector from the origin to the antenna element $\boldsymbol{A}_{nm}$, $\boldsymbol{S}(\psi_k, \alpha_k)$ is the signal vector, and the $(\cdot)$ operations is the dot product between two vectors. The coordinate of $\boldsymbol{A}_{nm}$ can be expressed as (nd, md, 0), where d is the distance between adjacent antennas, therefore, $\boldsymbol{A}_{nm}$ can be expressed as:

$$\boldsymbol{A}_{nm} = [nd, md, 0]_T \tag{5}$$

where (T) is transpose operation of the vector. Similarly, the signal $\boldsymbol{S}(\psi_k, \alpha_k)$ from the azimuthal angle ψ_k and the elevation angle α_k can be expressed as:

$$\frac{S(\psi_k, \alpha_k)}{||S(\psi_k, \alpha_k)||} = [cos(\alpha_k)cos(\psi_k), sin(\alpha_k), cos(\alpha_k)sin(\psi_k)]^T \tag{6}$$

Combining all the above formula into Eq. 1, we can get the phase shift between antenna A_{00} and A_{nm}:

$$\Phi_{n,m}(\psi_k, \alpha_k) = \gamma_k e^{-j\frac{2\pi(ndcos(\alpha_k)cos(\psi_k)) + mdsin(\alpha_k))}{\lambda}} \tag{7}$$

That is, ψ_k and α_k will introduce a specific phase shift at different antenna. Suppose the size of antenna array is $N \times M$, and take the antenna in the origin as reference, the phase shift between each antenna and reference antenna can be write as:

$$\varPhi_(\psi_k, \alpha_k) = \begin{bmatrix} 1 & \cdots & \varPhi_{0,M-1}(\psi_k, \alpha_k) \\ \vdots & \ddots & \vdots \\ \varPhi_{N-1,0}(\psi_k, \alpha_k) & \cdots & \varPhi_{N-1,M-1}(\psi_k, \alpha_k) \end{bmatrix} \tag{8}$$

the receiving signal due to kth path can be expressed as $\boldsymbol{a}(\psi_k, \alpha_k)$, where denotes the complex attenuation at the antenna in the origin along the path and $\boldsymbol{a}(\psi_k, \alpha_k)$ is a vector accumulated elements in the matrix by column, it can be expressed as:

$$\begin{aligned} \boldsymbol{a}(\psi_k, \alpha_k) = {} & [1 ... \varPhi_{N-1,0}(\psi_k, \alpha_k)\varPhi_{0,1}(\psi_k, \alpha_k) ... \varPhi_{N-1,1} \\ & (\psi_k, \alpha_k) ... \varPhi_{0,M-1}(\psi_k, \alpha_k) ... \varPhi_{N-1,M-1}(\psi_k, \alpha_k)]^T \end{aligned} \tag{9}$$

The vector $\boldsymbol{a}(\psi_k, \alpha_k)$ is called steering vector which represents the phase shift between different antennas theoretically. Because there are multiple propagation paths, we have multiple steering vectors. The overall steering matrix A is defined as:

$$A = [\boldsymbol{a}(\psi_1, \alpha_1), \boldsymbol{a}(\psi_2, \alpha_2), \boldsymbol{a}(\psi_L, \alpha_L)] \tag{10}$$

L represents the number of propagation path and the dimensions of A is $(N \times M) \times L$. The receiver receives a linear combination of the multiple reflections from different path, so the received signal can be expressed as:

$$\boldsymbol{x} = A\boldsymbol{\varGamma} \tag{11}$$

where A is the steering matrix and $\boldsymbol{\varGamma} = [\gamma_1, \gamma_1, ..., \gamma_L]^T$ represents the complex attenuations along L propagation paths. The standard MUSIC algorithm can be used for one-dimensional angle estimation, but it still applies to two-dimensional case.

In our scenario, when we get the vector X through experimental measurements, we can use the MUSIC algorithm to get the steering matrix A, and then we can easily derive the azimuth and elevation angle. The key idea behind the MUSIC algorithm is that the eigenvector of $\boldsymbol{x}\boldsymbol{x}^H$ corresponds to the eigenvalue zero represents noise subspace, If they exist, then they are orthogonal to the steering vector A which represents signal subspace. For simplicity, we omitted the formula deduction process, and if you are interested, you can refer to [].

However, directly using the above-mentioned measured vector $\boldsymbol{x}$ does not give a good result. It is theoretically proved that in order to obtain an eigenvector corresponding eigenvalue is zero of the matrix $\boldsymbol{x}\boldsymbol{x}^H$, the measured vector should be a matrix whose rows and columns are both larger than the number of multipaths [31]. A straightforward method is to use multiple measurements/packets to form a measurement matrix X, of which each column represents the result of a single measurement. However, in this paper, we want to observe the influence of the human activity on multiple packets, so we proposed an idea to obtain

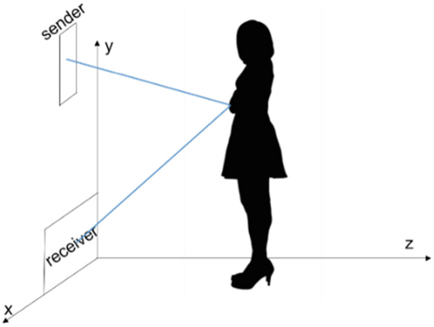


Fig. 4. Reflection off human body

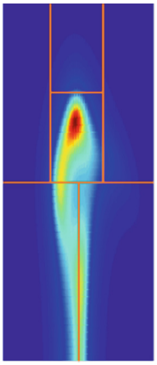


Fig. 5. Body segmentation

an imaging picture using only one data packet. We can increase the number of physical antennas to increase the accuracy. But building a physical array is very expensive and not suitable for real situations.

OFDM uses multiple subcarriers to transmit information, and the frequency of each subcarrier is different. Since the frequency interval of the subcarriers is small, the phase shift generated between the subcarriers is negligible for signals in a certain direction, which means that the steering matrix of different subcarriers is the same. In order to distinguish the phases of different subcarriers, the Tof (Time of Flight) is introduced and the phases of different subcarriers can be expressed as formula 12 [31]:

$$\varOmega(\tau_k) = e^{-j2\times\pi\times f_\delta\times\tau_k} \tag{12}$$

Finally, the steering vectors of different subcarriers of different antennas can be expressed as the kron product of formula 8 and formula 6.

Then, we can follow the classic MUSIC algorithm to solve the problem.

In one packet transmission, we can get the phase shift across different subcarriers of different antennas, For example, we use a 5300 NIC that can report the CSI of 30 subcarriers. We can get the following measurement matrix:

$$Xmatrix = \begin{bmatrix} csi_{0,0,1} & \cdots & csi_{0,0,30} \\ \vdots & \ddots & \vdots \\ csi_{N-1,M-1,1} & \cdots & csi_{N-1,M-1,30} \end{bmatrix} \tag{13}$$

Finally, transform X0 into one-dimentional column vector X:

$$Xmatrix = [csi_{0,0,1} \cdots csi_{N-1,M-1,30}] \tag{14}$$

With the above measurement matrix, the following algorithm can be used to get the final imaging results.

Algorithm summary:

1. Construct sample covariance matrix $R = \frac{1}{P}\sum_{i=1}^{P} XX^H$, where P is the number of subcarriers.

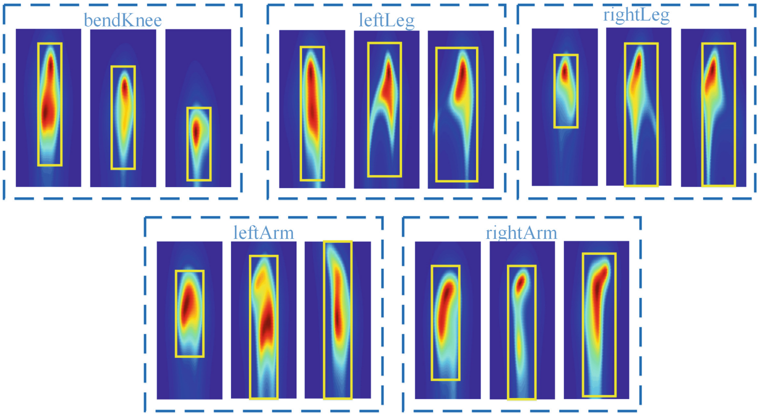


Fig. 6. The bounding box of human body changing as moving

2. Perform eigenvalue decomposition of the matrix R. Order eigenvectors of R according to eigenvalues. Let eigenvectors corresponding to L largest eigenvalues span signal subspace S, and remaining eigenvectors span noise subspace G.
3. Construct spatial spectrum

$Pmusic(\psi,\alpha)=\frac{1}{\boldsymbol{a}(\psi,\alpha)^{H}GG^{H}\boldsymbol{a}(\psi,\alpha)}$

Through the above steps, we can get the spatial spectrum Pmusic, which represents the possibility of the existence of a signal in each direction. Pmusic can be understood as the intensity of the signal in each direction called heat map.

4.2 Continuous Human Activity Segmentation

When human performs continuous activity, a set of image sequences can be obtained according to the imaging algorithm mentioned above, and many existing image processing algorithms can be used for continuous action segmentation. In this paper, a simple algorithm is proposed to verify the feasibility of continuous activity segmentation based on our heat map.

CHAR uses the body's reflection signal to measure the angle of each part. However at some point, our receiving antenna can only receive reflection from only some parts of the body. As the Fig. 4 shows, because the propagation of the signal satisfies the law of reflection, most parts of body's reflected signals can't be received by the receiver. However, because the chest is large and convex, its reflection signal is always the strongest. As shown in Fig. 5, we confirm the center of the image according to the strongest reflection position, and then divide a picture into the following six parts. The upper part of the chest represents the head, the left and right sides of the chest respectively represent the left and right arms, and the lower part of chest are the effect of the left and right legs respectively.

4.3 Human Activity Segmentation

A set of image sequences $p_1, p_2, ..., p_N$ can be obtained by using the imaging algorithm, and then we use the minimal area segmentation method to split the action [32]. First, the area of the bounding box is calculated by using the bounding box of the human body. The value of the area is used as an index to measure the degree of limb extension. The smaller the value, the closer the limb is to the body, and the larger the value, the greater the limb is stretching. The minimal value point is used as the action segmentation point. The key of the algorithm is to find the minimum value of the bounding box area function. In order to improve the noise resistance of the method and effectively locate the minimum value points, the smooth function is first executed. The result shows as the Fig. 6.

Smooth Body Bounding Box Area Function. Assume that Bt(x, y, w, h) is the minimum enclosing rectangle of the human body in the t-th frame, referred to as the human bounding box, where (x, y) represents the coordinates of the upper left corner of the human bounding box, and w and h respectively represent the surrounding width and height of the box. $S(t) = B_t^w \times B_t^h$ denotes the area function of the bounding box. The area of the human bounding box changes as the person moves.

In order to overcome the influence of the missing character extraction, find the essential regularity of the area function, we apply the local weighted smoothing method to smooth the are function, the steps are as follows:

1. Set the width of the local smoothing window to L. The smoothing target point is in the middle of the window. There are two neighbors on the left and right sides, and the localized weighted linear regression is performed on the target point. The regression model is $f(t) = \alpha_0 + \alpha_1 t$, where α_0 and α_1 are constant terms and primary coefficients, respectively. The performance indicator function is $J(\alpha_0, \alpha_1) = \frac{1}{L}\sum_{i=1}^{L} w_i(S_i - f(t_i))^2$, where S_i is the area value of the i-th point in the smoothing window, and the initial weight function is $w_i = (1 - |\frac{t-t_i}{d(t)}|^3)^3$, t is the target position, t_i is the i-th neighbor position of the t point in the smoothing window, and $d(t)$ is the farthest distance from the neighboring data point in the window.
2. Calculate the residual $r_i = S_i - f(t_i)$ of each data point in the window based on the weighted regression data.
3. Calculate the weight of each data point in the window, and define the weight as

$$w_i = \begin{cases} ((1 - (\frac{r_i}{6M})^2)^2 & r_i < 6M \\ 0 & r_i \geqslant 6M \end{cases} \tag{15}$$

   where r_i is the residual of the i-th data point, and M is the median of the absolute values of the L residuals, which is used to measure the degree of dispersion of the residual. If $r_i < 6M$, the corresponding weight is close to 1, if $r_i \geqslant 6M$, the weight is 0.

4. Re-execute the weighted linear regression function of step 1 and setting iterations being 5, using the regression model as the smoothing model.

After smoothing, most of the fluctuation points can be eliminated, making the area function more regular.

Action Segmentation. Let $S^{'}(t)$ be the area function after smoothing. If $t^{'}$ satisfies the inequality

$$(S^{'}(t^{'}+1)-S^{'}(t^{'}))\times(S^{'}(t^{'})-S^{'}(t^{'}-1))\leqslant 0 \tag{16}$$

Then, $S^{'}(t^{'})$ is an extreme value of the function $S^{'}(t)$

Considering the incompleteness of the action at the start and end point, and in order to reduce the impact of insufficient smoothing, the extreme points calculated by equation(16) are subjected to secondary filtering in the space-time domain:

1. The area of the start frame and the end frame of the video is added to the set of extreme points, and the maximum or minimum value is determined according to the trend of change of $S^{'}(t)$.
2. Each extreme point S_i is checked in turn. If the time interval between S_i and the adjacent extreme point S_{i-1} is less than the threshold T_t, and their area difference is less than T_s, Then S_i is regarded as the interference point. According to multiple experiments, T_t is set to 0.2 s (that is, 25 frames/s, 5 frames apart), and T_s is set to $0.1\times min(S_{i-1},S_i)$.

After filtering twice, the attribute value is judged by the extreme point, if $(S^{'}(t^{'}+1)-S^{'}(t^{'}))>0$ is satisfied, it is a minimum value point. After obtaining the minimum value point of the area function, the frame sequence between the extracted minimum value points is sequentially divided into separate action segments, thereby achieving action segmentation.

4.4 IC^2-Based Activity Recognition

CHAR can obtain image sequences by using Wi-Fi imaging method. The framework of our classifier is showed in Fig. 7. The above-mentioned continuous motion segmentation algorithm can divide the sequence of pictures into several sequences according to the actions performed. Each sequence represents a complete action and we call it a sample. We process the input data through STN (Spatial Transformer Network) before the VGG19.

In traditional image classification, input data is a serials of samples which each sample is a three channel colored picture. Our input data is a set of consecutive action sequences represented in heat map format. Each sample is a 51 * 61 * 16 pixel picture which 51 being width, 61 being height and the number of each action being 16. In order to match the input channel of pictures, we transform the dimensionality to 272 * 61 * 3 while holding the amount of pixels unchanged.

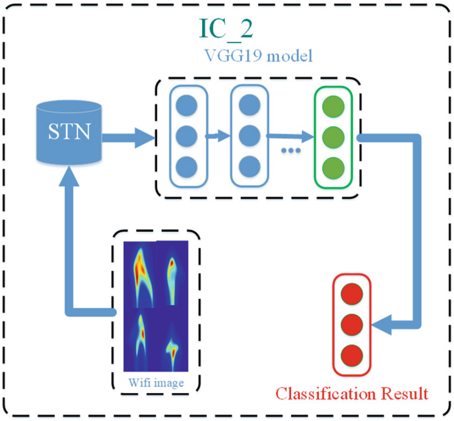


Fig. 7. Framework of IC^2

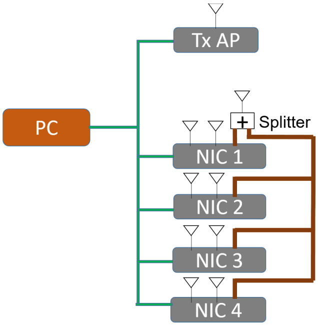


Fig. 8. System overview

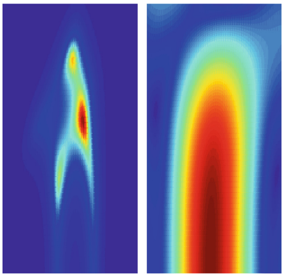


Fig. 9. Imaging result (left: CHAR, right: Wision)

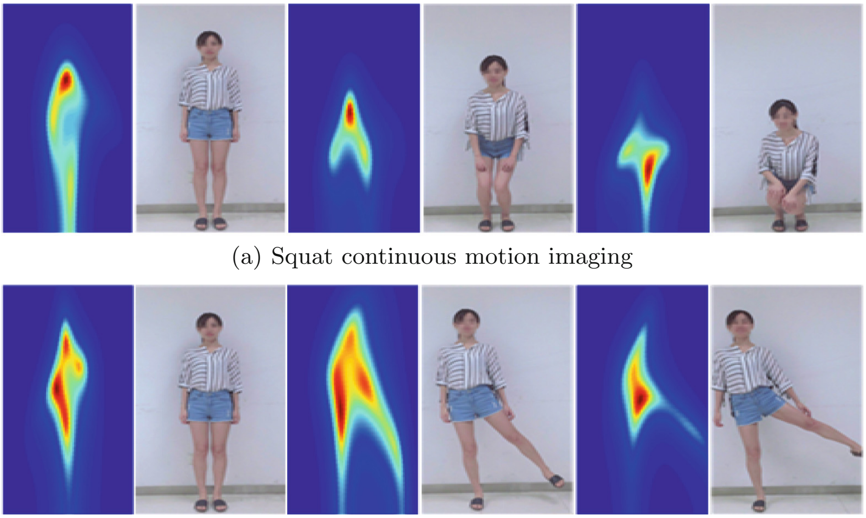
5 Implementation

We implemented our system using off-the-shelf Intel 5300 Wi-Fi NICs. We employed Linux CSI tool [68] to obtain the PHY layer CSI information for each packet. Our transmitter is directional antenna on the NIC, whose model is SCWL-2425-15D65VHPB-001. Its horizontal lobe width is 20° and the vertical lobe width is 70°. The object stands at a position two meters away from the antenna, and the beam of the antenna can cover the whole body of the person. Therefore, the use of directional antennas can effectively eliminate the effects of other objects.

Our receiver is a two-dimensional antenna array, the size of which is 4 × 4 using eight NICs. Because we use the phase difference between antennas to calculate the direction of arrival of the signal, different antennas should be synchronized. However, due to hardware errors, the antenna between different NIC has CFO, SFO, PDD and so on. We use the method proposed in Phaser[d] to calibrate the phase. Since the clock sources of different NIC are different, it is difficult to be calibrated. We send the signal of one antenna of one NIC to the other through the power splitter. We use the data of the sacrificed antenna to calibrate the phase between different NIC. So if we use 4 network cards can form a 3 × 3 receiving antenna array as shown in Fig. 8. The phase difference between different antennas of the same NIC can be calibrated by software. For more detailed principles, please refer to Phaser. In the following evaluation, we use 8 network cards to form a 4 × 4 receiving antenna array, in which one antenna data is not used.

6 Evaluation

We evaluate our prototype in an office building. First, CHAR uses a 2-D antenna array to evaluate the ability to image objects. Next, CHAR demonstrates the ability to identify different human activities using imaging results.



(a) Squat continuous motion imaging

(b) Left leg side lift continuous motion imaging

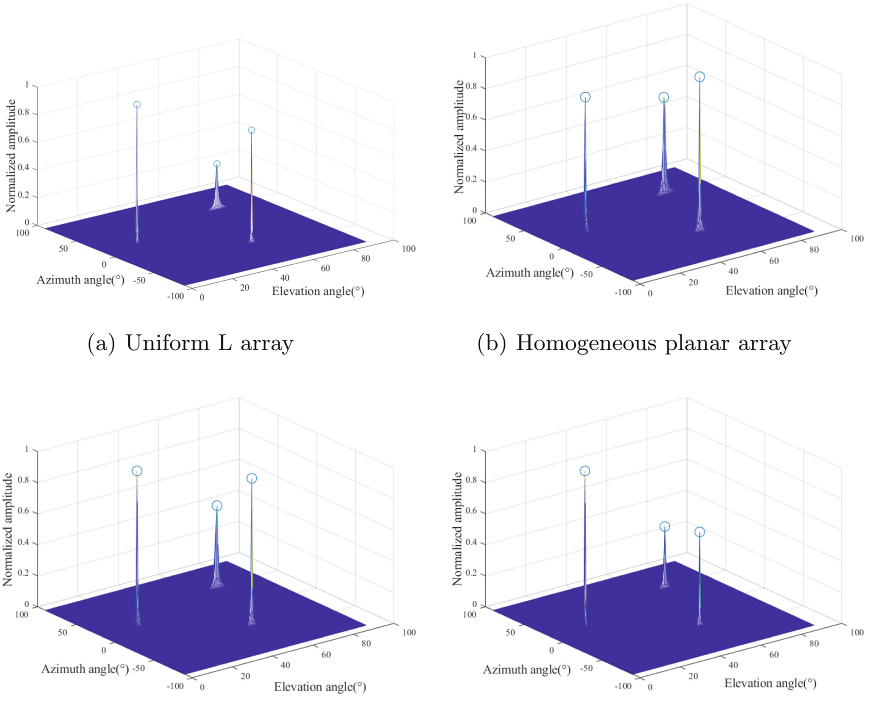
Fig. 10. Human figures obtained with CHAR

6.1 Imaging Using 2D Antenna Arrays

According to the analysis of 4.2, the reflected signal propagation conforms to the law of reflection. So in order to obtain the reflection signal of more body parts, we use two directional antennas as transmitting, which are placed at coordinates (10, 70) and (10, 140) respectively. We use 4×4 antenna array as receiver whose coordinate is (0, 0) and the distance between every two adjacent antennas is half wavelength. CHAR sends OFDM symbols which contain multiple subcarriers and the central frequency is 2.4 G.

Experimenter stands at a location of two meters away from the receiving antenna. In order to receive signal from different body part, experimenter should make a slight movement in situ, collects the data of two seconds, and achieve WiFi imaging using the algorithm proposed above in which multiple data packets are used for better imaging results. We compared the imaging results of CHAR and Wision [30]. In the specific implementation, the two systems sent the same data, and the imaging results are shown in Fig. 9.

Result: Due to the movement of the experimenter, different body parts will introduce a reflection signal, and the imaging result are shown in Fig. 10. We can see that the strongest reflection area is located in the chest part, and the reflection of the head and limbs is weak, but Wision's resolution is very low and in which different parts of the body can not be distinguished.



(a) Uniform L array

(b) Homogeneous planar array

(c) Uniform cross array

(d) Uniform circle array

Fig. 11. Comparison of different antenna arrays

6.2 Imaging Human Activity

To evaluate the human activity imaging performance of CHAR, we design five different actions namely left hand raising, right hand raising, left leg lifting, right leg lifting and squating. We have 5 participants and everyone preforms the five actions above for ten times.

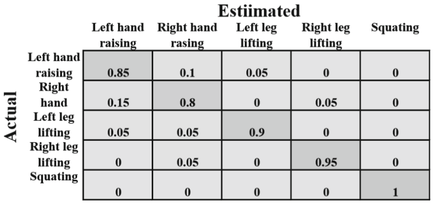
In order to see the imaging results of continuous motion, we used the information of the subcarriers shown in 4.1 for imaging. We can see that the imaging result changes as object performs actions with different data packets. For each motion, we choose the results of some representative packets. As shown in Fig. 10, figures on the left shows the imaging results of our system and those on the right shows the actual actions of the user. For the squatting action, we can see that as experimenter moves down, the strongest reflection point keeps moving down; for left leg lifting, we can see the change of a leg raised to the side. Different changes can be observed for the five simple actions.

6.3 Comparison of Different Antenna Arrays

In Sect. 4.2.2, this article describes several arrays of different shapes that can be used for 2D DOA estimation. In order to study the performance and effects of various arrays, MATLAB was used in this section for simulation experiments. In the experiment, the number of signals to be tested is $D = 3$, the signal-to-noise ratio is $SNR = 10\,\mathrm{dB}$, the number of snapshots is N = 100, and the azimuth and elevation angles of the three sources are: $(-18°., 18°), (18°)., 27°), (46.8°, 57.6°)$. The two-dimensional spectrum search is performed in the range of azimuth angle $-90°$ to $90°$ and pitch angle 0 to $90°$, and the angle search interval is $0.05°$. Except for a uniform circular array, the distance between adjacent antennas is $\lambda/2$, and the radius of the circular array is λ. Using this distance can effectively resist the phase ambiguity problem, and the specific principle is beyond the scope of this paper. The circular array has 8 array elements, and the plane array, cross array and L array have 9 array elements. The 2D MUISC results for the four different arrays are shown in Fig. 11.

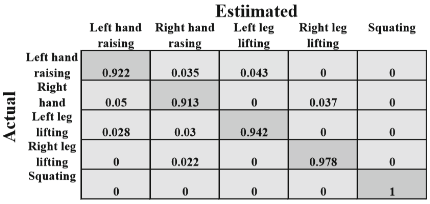
The X axis represents the azimuth angle, the Y axis represents the pitch angle, and the Z axis represents the magnitude of the MUSIC spectrum obtained. It can be understood as the signal strength of the angle, and the circle represents the estimated angle information. Comparing the four graphs, it can be found that in the case where the number of antenna elements is similar, the spectrum of the uniform circular array and the uniform cross array is sharper, and the plurality of spectral peaks are relatively uniform, indicating that its angle measuring ability is stronger. However, in reality, the uniform planar array has a smaller aperture and a smaller footprint, and the theoretical model is closer to the real scene. Therefore, a uniform planar array is used in the actual experiment.

Overall, we have sufficient resolution for our imaging systems to meet imaging requirements. Although the uniform L-array performs well, the angular accuracy and stability are very strong, but its array aperture is large, and the actual area occupied is large. Therefore, it is quite different from the signal propagation



		Estiimated				
		Left hand raising	Right hand rasing	Left leg lifting	Right leg lifting	Squating
Actual	Left hand raising	0.85	0.1	0.05	0	0
	Right hand	0.15	0.8	0	0.05	0
	Left leg lifting	0.05	0.05	0.9	0	0
	Right leg lifting	0	0.05	0	0.95	0
	Squating	0	0	0	0	1

Fig. 12. Confusion matrix of activity classification with SVM



		Estiimated				
		Left hand raising	Right hand rasing	Left leg lifting	Right leg lifting	Squating
Actual	Left hand raising	0.922	0.035	0.043	0	0
	Right hand	0.05	0.913	0	0.037	0
	Left leg lifting	0.028	0.03	0.942	0	0
	Right leg lifting	0	0.022	0	0.978	0
	Squating	0	0	0	0	1

Fig. 13. Confusion matrix of activity classification with IC^2

model and is not suitable for the actual scene. Uniform planar arrays have poor overall performance. Uniform circular arrays and uniform cross arrays have good direction finding accuracy and stability, and can perform two-dimensional direction finding on multiple incoherent sources.

6.4 Activity Recognition

We test our data in five different actions and each action contains 500 samples. 80% are used as training sets and 20% test sets. The parameters for IC^2 are set as follows, learning rate 0.01, epoch 100 and batch 75.

Confused Matrix Comparison. The confused matrix shows both SVM and IC^2 can get at least 80% classification accuracy. As Figs. 12 and 13 show, especially in squat moving which can capture more representative features than other moving actions. Accuracy can reach up to 1 and no misclassification. We analyze the statistics through comparing the squat moving with other moving actions, the previous action can track the reflected signal from chest up and down with moving which can lead to strong representative features. The worst case is classifying the right hand raising action which the accuracy is 80%. Because classification is based on continues sequences partition, arm and leg can not reflect strong signals due to their physical shapes.All CNN based classification approaches are beyond 91%.

Classification Accuracy Comparison. Both SVM and IC^2 can reach up to very high accuracy when the number of classification is small. Figure 14 shows that as the number increases, IC^2 begins to show more advantages than SVM. The average accuracy of IC^2 still maintains in high level even the difficulty increased. During our test, We observe that IC^2 network is much more robust than SVM. For different test samples, once the loss is converged in training data set, the evaluation accuracy will always maintain in very high level rather than SVM that has very fluctuation.

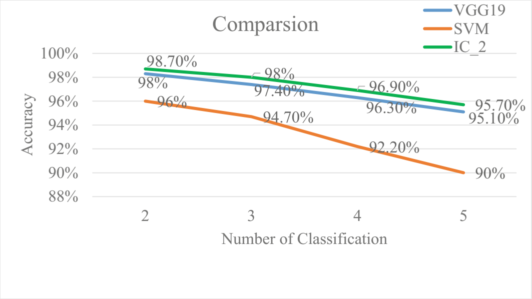


Fig. 14. Accuracy comparison between SVM, VGG19 and IC^2

7 Conclusion

Indoor wireless sensing has spawned numerous applications in a wide range of living, production, commerce, and public services. The increase of mobile and pervasive computing has sharpened the need for accurate, robust, and off-the-shelf indoor continuous action recognition schemes. CHAR can easily solve automatic segmentation and action recognition problem using WiFi imaging which is achieved using the transmitted signals reflected from different body parts. We propose a novel approach using these reflections to realize Wi-Fi imaging. The evaluations demonstrate that CHAR can reach an average 95% high matching accuracy under a wide variety of environment.

References

1. Harville, M., Li, D.: Fast, integrated person tracking and activity recognition with plan-view templates from a single stereo camera. In: Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2004, vol. 2, p. II. IEEE (2004)
2. Fullwood, D., Kalidindi, S., Adams, B., Ahmadi, S.: A discrete fourier transform framework for localization relations. Comput. Mater. Continua (CMC) **9**(1), 25 (2009)
3. Kwapisz, J.R., Weiss, G.M., Moore, S.A.: Activity recognition using cell phone accelerometers. ACM SigKDD Explor. Newsl. **12**(2), 74–82 (2011)
4. Pu, Q., Gupta, S., Gollakota, S., Patel, S.: Gesture recognition using wireless signals. GetMobile: Mob. Comput. Commun. **18**(4), 15–18 (2015)
5. Liu, W., Luo, X., Liu, Y., Liu, J., Liu, M., Shi, Y.Q.: Localization algorithm of indoor wi-fi access points based on signal strength relative relationship and region division. Comput. Mater. Continua **55**(1), 071–071 (2018)
6. Wang, Y., Liu, J., Chen, Y., Gruteser, M., Yang, J., Liu, H.: E-eyes: device-free location-oriented activity identification using fine-grained wifi signatures. In: Proceedings of the 20th Annual International Conference on Mobile Computing and Networking, pp. 617–628. ACM (2014)
7. Wang, G., Zou, Y., Zhou, Z., Wu, K., Ni, L.M.: We can hear you with Wi-Fi!. IEEE Trans. Mob. Comput. **15**(11), 2907–2920 (2016)

8. Wang, Y., Wu, K., Ni, L.M.: WiFall: device-free fall detection by wireless networks. IEEE Trans. Mob. Comput. **16**(2), 581–594 (2017)
9. Wang, W., Liu, A.X., Shahzad, M., Ling, K., Lu, S.: Understanding and modeling of wifi signal based human activity recognition. In: Proceedings of the 21st Annual International Conference on Mobile Computing and Networking, pp. 65–76. ACM (2015)
10. Xi, W., et al.: Device-free human activity recognition using CSI. In: Proceedings of the 1st Workshop on Context Sensing and Activity Recognition, pp. 31–36. ACM (2015)
11. Yang, L., Chen, Y., Li, X.-Y., Xiao, C., Li, M., Liu, Y.: Tagoram: real-time tracking of mobile RFID tags to high precision using COTS devices. In: Proceedings of the 20th Annual International Conference on Mobile Computing and Networking, pp. 237–248. ACM (2014)
12. Ding, H., et al.: Device-free detection of approach and departure behaviors using backscatter communication. In: Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing, pp. 167–177. ACM (2016)
13. Zhao, Y., Patwari, N., Phillips, J.M., Venkatasubramanian, S.: Radio tomographic imaging and tracking of stationary and moving people via kernel distance. In: 2013 ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN), pp. 229–240. IEEE (2013)
14. Wilson, J., Patwari, N.: Radio tomographic imaging with wireless networks. IEEE Trans. Mob. Comput. **9**(5), 621–632 (2010)
15. Zhao, Y., Patwari, N.: Noise reduction for variance-based device-free localization and tracking. In: 2011 8th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks, pp. 179–187. IEEE (2011)
16. Han, J., Qian, C., Yang, P., Ma, D., Jiang, Z., Xi, W., Zhao, J.: GenePrint: generic and accurate physical-layer identification for UHF RFID tags. IEEE/ACM Trans. Netw. **24**(2), 846–858 (2016)
17. Sigg, S., Blanke, U., Tröster, G.: The telepathic phone: frictionless activity recognition from WiFi-RSSI. In: 2014 IEEE International Conference on Pervasive Computing and Communications (PerCom), pp. 148–155. IEEE (2014)
18. Sigg, S., Shi, S., Buesching, F., Ji, Y., Wolf, L.: Leveraging RF-channel fluctuation for activity recognition: active and passive systems, continuous and RSSI-based signal features. In: Proceedings of International Conference on Advances in Mobile Computing and Multimedia, p. 43. ACM (2013)
19. Xi, W., et al.: Electronic frog eye: counting crowd using WiFi. In: IEEE INFOCOM 2014-IEEE Conference on Computer Communications, pp. 361–369. IEEE (2014)
20. Zhou, Z., Yang, Z., Wu, C., Shangguan, L., Liu, Y.: Towards omnidirectional passive human detection. In: Proceedings IEEE INFOCOM, pp. 3057–3065. IEEE (2013)
21. Kotsiantis, S.B., Zaharakis, I., Pintelas, P.: Supervised machine learning: a review of classification techniques. Emerg. Artif. Intell. Appl. Comput. Eng. **160**, 3–24 (2007)
22. Kim, Y.: Convolutional neural networks for sentence classification. arXiv preprint arXiv:1408.5882 (2014)
23. Girshick, R.: Fast R-CNN. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1440–1448 (2015)
24. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)

25. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
26. Jaderberg, M., Simonyan, K., Zisserman, A., et al.: Spatial transformer networks. In: Advances in neural Information Processing Systems, pp. 2017–2025 (2015)
27. Woo, S., Park, J., Lee, J.-Y., So Kweon, I.: CBAM: convolutional block attention module. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 3–19 (2018)
28. Shen, W.-L., Lin, K.C.-J., Gollakota, S., Chen, M.-S.: Rate adaptation for 802.11 multiuser mimo networks. IEEE Trans. Mob. Comput. **13**(1), 35–47 (2014)
29. Halperin, D., Hu, W., Sheth, A., Wetherall, D.: Predictable 802.11 packet delivery from wireless channel measurements. ACM SIGCOMM Comput. Commun. Rev. **41**(4), 159–170 (2011)
30. Huang, D., Nandakumar, R., Gollakota, S.: Feasibility and limits of Wi-Fi imaging. In: Proceedings of the 12th ACM Conference on Embedded Network Sensor Systems, pp. 266–279. ACM (2014)
31. Kotaru, M., Joshi, K., Bharadia, D., Katti, S.: SpotFi: decimeter level localization using WiFi. In: ACM SIGCOMM Computer Communication Review, vol. 45, no. 4, pp. 269–282. ACM (2015)
32. Wang, L., Suter, D.: Learning and matching of dynamic shape manifolds for human action recognition. IEEE Trans. Image Process. **16**(6), 1646–1661 (2007)