Robust and Effective Component-Based Banknote Recognition for the Blind

Faiz M. Hasanuzzaman, Xiaodong Yang, and YingLi Tian, *Senior Member, IEEE*

Abstract—We develop a novel camera-based computer vision technology to automatically recognize banknotes to assist visually impaired people. Our banknote recognition system is robust and effective with the following features: 1) high accuracy: high true recognition rate and low false recognition rate; 2) robustness: handles a variety of currency designs and bills in various conditions; 3) high efficiency: recognizes banknotes quickly; and 4) ease of use: helps blind users to aim the target for image capture. To make the system robust to a variety of conditions including occlusion, rotation, scaling, cluttered background, illumination change, viewpoint variation, and worn or wrinkled bills, we propose a component-based framework by using speeded up robust features (SURF). Furthermore, we employ the spatial relationship of matched SURF features to detect if there is a bill in the camera view. This process largely alleviates false recognition and can guide the user to correctly aim at the bill to be recognized. The robustness and generalizability of the proposed system are evaluated on a dataset including both positive images (with U.S. banknotes) and negative images (no U.S. banknotes) collected under a variety of conditions. The proposed algorithm achieves 100% true recognition rate and 0% false recognition rate. Our banknote recognition system is also tested by blind users.

Index Terms—Banknote recognition, blind and visually impaired, component based, computer vision, speeded up robust feature (SURF).

I. INTRODUCTION

WORLD Health Organization approximates that there were 161 million visually impaired people around the world in 2002, which is about 2.6% of the total population. Among these statistics, 124 million had low vision and 37 million were blind [15]. Visually impaired people face a number of challenges when interacting with the environments because so much information is encoded visually in daily life. One specific difficulty that a blind person would encounter is to know the value of the currency or bill he or she is holding. Currently, printed denominations of U.S. currency are $1, $2, $5, $10, $20, $50, and $100. With a few recent exceptions, all of the banknotes are identical in size and color, and inaccessible to people who are blind or significantly visually impaired. Although a federal judge has ruled that the U.S. Treasury Department is violating the law by failing to design and issue currency that is readily distinguishable to blind and visually impaired people, in November 2006, the Treasury Department argued that any change would be costly, with estimates ranging from $75 million for equipment and $9 million in annual expenses to punch holes, to $178 million in one-time charges and $50 million in annual expenses to print different-size bills [20]. Some latest redesigned money has been issued with an enlarged color number for visually impaired people. However, it may take years to issue bills with additional blind-friendly changes.

According to the American Foundation for the Blind [1], one way that a blind person can identify paper currency is to fold each denomination in different ways. The recommendation for folding some currency is to fold $5 bills lengthwise, $10 bills widthwise, $20 bills are folded twice, and $1 bills may remain unfolded and put into a wallet. Although the idea of folding the bills is good, it needs others' help to organize bills. Recent advances in theories, sensors, and embedded computing hold the promise to enable computer vision technique to address their needs. Although a number of studies on camera-based bill recognition have been published in pieces of literature [3], [9], [10], [13], [16]–[18], [31]–[33], some of them are restricted to specific and standard environment. For example, the whole bill must be visible without occlusion, wrinkles, etc. Furthermore, none of the existing methods provides the analysis of recognition of negative images (i.e., images do not contain banknotes) and aiming guidance for blind users. An automatic banknote recognition system for blind people should be able to recognize banknotes in a wide variety of real-world environments, such as occlusions, cluttered background, varied illumination, different viewpoints, and worn or wrinkled bills, as well as to provide the feedback of aiming guidance.

The extraction of sufficient, stable, and distinctive monetary features is significant for accuracy and robustness of a banknote recognition algorithm. Recent developments of interest points detectors and descriptors in computer vision, such as scale invariant feature transform (SIFT) [19] and speeded up robust feature (SURF) [2], enable the extraction of sufficient, stable, and distinctive recognition of monetary characteristics by using local image features. The sophisticated designs of local image features make them robust to geometric and photometric variations, such as rotation, scaling, viewpoint variation, and illumination change. These properties make them well adapted to detect objects appearing with different scales and orientations in images.
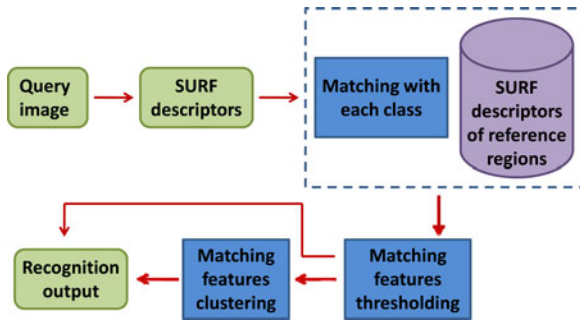
Fig. 1.　System diagram of the proposed banknote recognition for the blind.

In this paper, we propose a novel component-based banknote recognition system by using SURF features to achieve high recognition accuracy and to handle various challenging conditions in real-world environments. Fig. 1 shows our system diagram. First, monetary features of each query image are extracted by SURF. These features are then matched with the precomputed SURF features of reference regions of the ground truth image in each banknote category. The numbers of matched features are compared with automatic thresholds of each reference region to determine the banknote category. Furthermore, the spatial relationship of matched features is employed to avoid false recognition with negative images and provide aiming guidance to blind users for bill image capture. Then, the system outputs the recognition result.

An earlier version of this paper can be found in [10]. Compared with our previous work, there are three major extensions that merit being highlighted: 1) the previous work manually set the matching thresholds for each region. The automatic thresholding method proposed in this paper can learn all the thresholds from training images so that the proposed method can be easily extended to recognize banknotes from different countries (see Section III-C); 2) the previous work was not able to handle negative images (i.e., query images containing no bills). In this paper, we propose a spatial clustering method to detect if there is a bill in the camera view and further use the detection result to guide the blind users to correctly aim the target (see Section III-D). This process significantly reduces false recognition rate; and 3) we enlarge the evaluation dataset and add more experimental results and demonstrate the efficiency and robustness of the proposed work. In addition, the proposed system is evaluated by nine blind users. Overall, the work introduced in this paper offers the following main contributions to robust and efficient recognition of banknotes in a wide variety of real-world environments for the blind.

1) We propose a component-based framework to handle a variety of conditions including occlusion, cluttered background, viewpoint variation, and worn or wrinkled bills due to the following reasons: a) components retain more discriminative information for banknote categories, while other parts of the banknotes are similar across different categories; b) components vary much less than the global pattern under the geometric transform; and c) the component-based model is insensitive to partial occlusions and background clutter.

2) We propose a new automatic method to obtain matching threshold (see Section III-C). This benefits the extension of our system to recognize banknotes of different countries.

3) We employ SURF features in the banknote recognition system to handle rotation, scaling, and illumination changes due to their robustness for geometric and photometric variations, and great speed improvement.

4) We propose a new schema by employing the spatial information of matched SURF features to avoid false recognitions (i.e., recognize negative images as banknotes).

5) To solve the aiming problem of blind users, we further provide audio guidance to the user if there is no banknote in the camera view based on the recognition results.

6) The proposed banknote recognition system is evaluated by nine blind subjects.

II. RELATED WORK

Over the past 50 years, there have been many efforts to develop electronic aids to improve quality of life and the safety of individuals with special needs [4], [6], [21]–[23], [25], [29], [30], [34]–[36], [38], [39], [45]–[47]. Several techniques have been developed to identify banknotes from camera captured images to help blind or visually impaired people. Lee *et al.* [18] proposed a method to extract features from specific parts of Euro banknotes representing the same color. In order to recognize banknotes, they used two key properties of banknotes: direction (front, rotated front, back, and rotated back) and face value (5, 10, 20, 50, 100, 200, and 500). They trained five neural networks for insert direction detection and face value classification. The results showed a high recognition rate (about 98.5%) and a low training period. Frosini *et al.* also presented a neural-network-based bill recognition and verification method [5]. The learning vector quantization method to recognize eight kinds of Italian Liras is proposed in [16] and [17]. Most banknote recognition methods employed neural network techniques for classification [13], [16]–[18], [31]–[33].

Reiff and Sincak [27] used SIFT detector and descriptors to classify Slovak banknotes in a well-controlled environment. Symmetrical masks have been used by Vila *et al.* to consider specific signs in a paper currency [37]. In their method, the summation of nonmasked pixel values in each banknote is computed and fed to a neural network. This method considers images of both the front and back of the paper currency, but only the front image is used for recognition. In the approach in [40], the patterns of an edge on a banknote are used for recognition, and the image of a banknote is vertically divided into a number of equal small parts. Then, the number of pixels associated with edges detected in each part are counted and fed to a three-layer backpropagation neural network for recognition. Hassanpour and Farahabadi [9] proposed a hidden Markov model (HMM)-based method. By employing the HMM, the texture characteristics of paper currencies are modeled as random processes and can be extended to distinguish paper currency from different countries. In the aforementioned methods, the whole bill must be visible without occlusion, wrinkles, etc. Neither of these methods

Fig. 2. Portable bill recognition product for blind users in the market "Note Teller 2" [26].

analyzes the false recognition nor considers the aiming problem of blind users.

As shown in Fig. 2, a portable bill recognition product in the market is called "Note Teller 2," which is manufactured by BRYTECH, Ottawa, ON, Canada [26]. The bill must be inserted in a slot on the top of this device. Based on the evaluation of AFB TECH [11], the overall identification accuracy of "Note Teller 2" is about 80% and it has difficulty in identifying worn or wrinkled bills. Compared with the constrained recognition environment of "Note Teller 2," our system enables banknote recognition in an open environment, i.e., blind users only need to hold a bill and take an image of the bill (e.g., by a wearable camera on sunglasses or a cell phone camera).

Similar to banknote recognition, systems for coin recognition have also been proposed. Huber *et al*. [41] applied a multistage approach to coin recognition. In this work, segmentation and rotational angle estimation were used to achieve translation and rotation invariance. Principal components analysis was further employed to tackle illumination change. Shen *et al*. [42] proposed to extract local texture features by Gabor wavelets and local binary pattern to represent coin images. Because of experimental settings and properties of coins, the aforementioned coin systems did not take into account of the conditions, such as partial occlusion, cluttered background, scaling change, and wrinkling, which, however, are quite common in the scenario of banknote recognition for the blind people.

III. Proposed Method

A. Component-Based Framework for Banknote Recognition

The proposed banknote recognition algorithm is based on a component-based model. It has four main advantages over the global model: 1) the class-specific information is not evenly distributed on the banknote. Some regions cover more obvious class-specific features, while other regions are relatively similar across different classes. It will be more effective to use those more class-specific components in the recognition of banknotes.

2) A component-based model is able to focus on local and stable parts, which vary much less than the pattern of an entire banknote under the geometric and photometric changes. 3) Local image features that are generated from components are much less than that from the entire image. This helps to speed up the matching process and reduce memory requirement. 4) A component-based model is more robust in handling partial occlusions. It is empirically impossible to take account of all conditions which cover the spectrum of possible variations that can result from occlusions. In the component-based model, individual components are detected by their corresponding detectors. Partial occlusions only affect the outputs of a portion of component detectors. As long as a certain amount of components are detected, the whole banknote is still able to be recognized.

Component Generation: In our banknote recognition system, ground truth images for both front side and back side of each category banknote are first collected under optimal conditions. Fig. 3 shows the reference banknote images of front and back sides of $1, $2, $5, $10, $20, $50, and $100 U.S. bills. The marked regions in red are the components that are chosen as reference regions for each category. For example, in the $5 bill, the distinguish-specific information in the front side is the number "5," the word "FIVE," and the face picture of Lincoln. Similarly, on the back side of the $5 bill, the number "5," the word "FIVE," and Lincoln Memorial are the most discriminative regions. Note that the golden number "10" on the right-bottom corner of the front side in $10 bill (also for the golden numbers "20" and "50" in $20 and $50 bills, respectively) is not taken into account because the matching of local image features within this region does not perform well. The same principal applies to all the other bills on both front and back sides.

In our system, a total of 14 images of seven categories of bills ($1, $2, $5, $10, $20, $50, and $100) with front and back sides are taken as ground truth images. The resolution of all the ground truth banknote images is 835×354. The size of each reference region is manually cropped. For each component, SURF features are extracted and saved for later matching with the query images in the recognition process.

B. Local Image Feature Detection and Representation

The detection and representation of image features are essential to many applications, such as image retrieval and recognition of object, texture, and scene categories, as they are more resistant to partial occlusion, background clutter, and viewpoint changes [7], [8], [14], [28]. This has motivated the development of several scale- and rotation-invariant local image detectors and descriptors. Generally, detectors localize interest points with stable regions in a scale space; descriptors build the representation of the support regions provided by detectors. The performance evaluations [2] of local image features observed SURF were comparable with and often better than state-of-the-art systems, such as SIFT [19] and gradient location and orientation histogram [24]. Furthermore, SURF can be computed and compared much faster, which is desirable in real-time applications. We choose SURF as the detection and representation of local image features based on the following reasons. 1) Banknote

Fig. 3. Reference regions with class-specific features on the bills of $1, $2, $5, $10, $20, $50, and $100. Each region marked in a red box is one reference region.

TABLE I
CONFUSION MATRIX TO MATCH RESULTS OF LOCAL IMAGE
FEATURES EXTRACTED BY SURF

	$1	$2	$5	$10	$20	$50	$100
$1	**1.00**	0.19	0.21	0.24	0.24	0.24	0.19
$2	0.18	**1.00**	0.23	0.20	0.23	0.25	0.15
$5	0.27	0.29	**1.00**	0.32	0.34	0.49	0.23
$10	0.24	0.22	0.26	**1.00**	0.33	0.36	0.21
$20	0.23	0.25	0.36	0.29	**1.00**	0.32	0.23
$50	0.17	0.19	0.31	0.27	0.29	**1.00**	0.17
$100	0.32	0.33	0.27	0.36	0.43	0.39	**1.00**
Negative	0.31	0.28	0.15	0.26	0.38	0.31	0.21

Each row corresponds to the averaged matching result of each banknote class and negative class to the reference regions.

space by upscaling the sizes of box filters rather than iteratively reducing the image size. The location of interest points in the image and over scale space is finally determined by the maximum determinant of approximated Hessian matrix in a $3 \times 3 \times 3$ neighborhood. In order to achieve rotation invariance, the SURF descriptor first identifies the dominant orientation of a support region by the Gaussian-weighted Haar wavelet responses. The support region is, then, oriented along its dominant orientation. The support region is quantized into 4×4 Cartesian location bins. Each location bin is further divided into 5×5 sample points. Haar wavelet responses at each sample point are computed including horizontal response d_x and vertical response d_y. The responses are further weighted by a Gaussian window to be robust to geometric disturbance and localization error. The sum $d_x, d_y, |d_x|$, and $|d_y|$ over 5×5 sample points constitute the feature vector for each location bin. Concatenating the feature vectors of 4×4 location bins generates the SURF descriptor with 64 dimensions. The descriptor is, then, normalized to a unit vector to be contrast invariant.

2) SURF Evaluation on Banknotes: In order to validate the repeatability, distinctiveness, and robustness of local image features extracted by SURF in banknote recognition, we evaluate the matching performance between the reference regions and banknote images taken under a variety of circumstances. We use 140 banknote images containing 20 images for each category of U.S. bill ($1, $2, $5, $10, $20, $50, and $100) and 50 negative images (see Fig. 7 for image examples). The banknote images are collected from a wide variety of conditions, such as occlusion, rotation, changes of scaling, illumination, and viewpoints. The negative images include both indoor and outdoor scenes. The reference regions of 14 ground truth images of seven categories of bills are used to match with training images based on SURF. If the distance between SURF descriptors of two interest points is below a certain threshold, the two interest points are matched.

The matching results are described in the confusion matrix in Table I. Each row in Table I corresponds to the average matching results for banknote images of each category. Element (i, j) corresponds to the average number of matched points between banknote images of category i and reference regions of category j. Each row is further normalized by the maximum number of matching points in this row. As we see, the best matching cases all occur at the diagonal elements of the confusion matrix. In addition, off-diagonal elements are much smaller than 1. This

images could be taken under the circumstances of rotation and scaling change. Interest points with support regions detected by SURF are robust to rotation and scaling change. 2) Descriptors associated with support regions are distinctive and compact for feature matching. 3) The low computational cost of SURF facilitates fast interest point localization and matching. In order to make this paper more self-contained, we briefly discuss the construction process of SURF in this section.

1) Interest Points Detection and Description: SURF is a combination of detector and descriptor which are computationally effective. The significant speed improvement of SURF detector is due to efficient use of integral images. The SURF detector is based on Hessian matrix [2] because of its good performance in detecting blob-like structures. The Hessian matrix is further approximated by using box filters to substitute for discretized second-order Gaussian derivatives. This enables the fast evaluation of second-order Gaussian partial derivatives by taking advantage of integral images. Thanks to the use of box filters and integral images, the SURF detector builds the scale
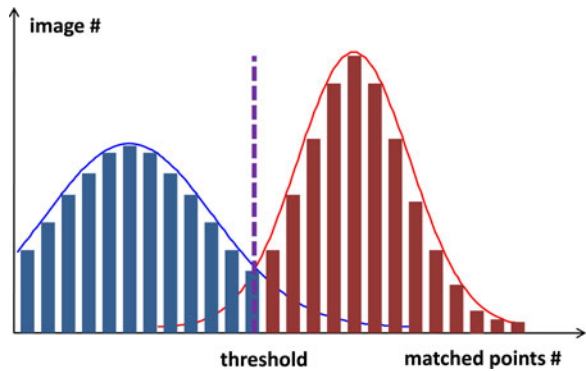
Fig. 4. Automatic threshold generation for each reference region in different banknote categories. The red distribution is the histogram of matched point # between a reference region in category i and training images containing the banknote belonging to category i. The blue distribution is the case between the reference region in category i and remaining images in the training set. The horizontal coordinate of the intersection point of the two distributions is used as the automatic threshold for this reference region.
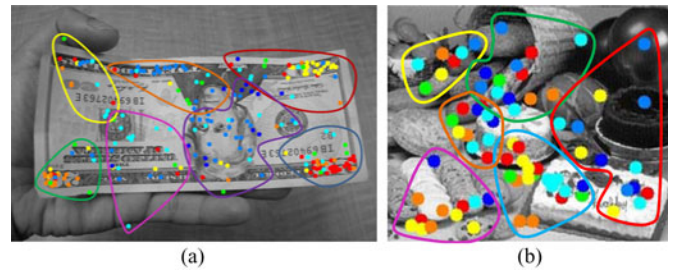

Fig. 5. Spatial relationship between matched points. Points matched to the same reference region are in the same color. Points grouped to the same cluster are enclosed by the curve with the same color. (a) In the positive image, the points in the same color tend to group in the same cluster for the positive image. (b) In the negative image, the matched points almost randomly distribute on the image for the negative image.

demonstrates that SURF descriptors are of great distinctiveness to distinguish local image features from different banknote categories and backgrounds. Therefore, it can be concluded from this confusion matrix that the SURF is able to represent and match local image features of banknotes accurately even under a variety of geometric and photometric variations.

C. Recognizing Banknotes by Automatic Feature Matching Thresholding

As shown in Fig. 1, given a query image, SURF first detects the interest points and generates corresponding descriptors. The precomputed SURF descriptors of reference regions in each category are then used to match with the extracted descriptors of the query image. The number of matched points between the query image and reference regions of different categories determines the reference region matching results. If N_i^j, the number of matched points between the query image and R_i^j (i.e., the reference region j in category i) is larger than T_i^j, the automatic threshold of R_i^j, then R_i^j, the reference region j in category i, is matched to this query image.

Here, we propose an automatic thresholding method to obtain T_i^j for each reference region by the histogram of the matched points numbers between reference regions and images in our training dataset. The training set includes 140 positive images with 20 images in each category and 50 negative images. For each R_i^j, its precomputed SURF descriptors are matched with extracted descriptors from training images of category i and the rest training images, respectively. The matching results are used to build the matching histogram, as shown in Fig. 4. The red distribution is the histogram of matched point numbers between R_i^j and training images of category i. Similarly, for the blue distribution, it describes the matching result between R_i^j and the rest images in the training set. The training images in category i contain R_i^j, while the rest images in the training set do not have such a region. Therefore, the average number of matched points of the red histogram is larger than that of the blue histogram. Due to partial occlusions, geometric and photometric variations of the images, the distributions of the matched points numbers present variances in both cases. The number of matched points corresponding to the intersection point between the two histograms is used as the automatic threshold T_i^j for R_i^j. The automatic thresholds of all the reference regions in different categories can be generated this way.

To recognize the category of a query image, all of the reference regions that are matched to the query image are retained. In the next step, the number of matched reference regions in each category is employed to determine the recognition category. The query image is recognized as category i if this image has the maximum number of matched reference regions with category i, and this number is not less than 2. From our recognition experiments, two matched reference regions are flexible in handling partial occlusions and are confident in recognizing the banknote category as well. If more than one category satisfies the aforementioned recognition rule, the total number of matched points in all matched reference regions is used to determine the final recognition category which maximizes the total number.

D. Spatial Clustering of Matching Features to Handle Negative Images and Aiming Problem of Blind Users

We discussed the recognition of a query image of a bill in previous sections. However, aiming problem is a common issue for blind users and will lead to many query images with no bill in the camera view (i.e., negative images). For some negative images with highly cluttered background, the number of matched points and the number of matched reference regions might pass the matching features thresholding rules (see Section III-C). In this case, a negative image will also be recognized as a certain banknote category, which is obviously a false recognition. Therefore, in order to assist the blind users in aiming at a banknote and to avoid false negative recognitions, we employ the spatial relationship between matched points to validate if a query image is really a bill. As shown in Fig. 5, the matched points matching to the same reference region are displayed in the same color. In the true recognition, which is shown in Fig. 5(a), matched points with the same color tend to cluster together as the spatial constraint of every matched region. While for the false case, which is shown in Fig. 5(b), matched points with the
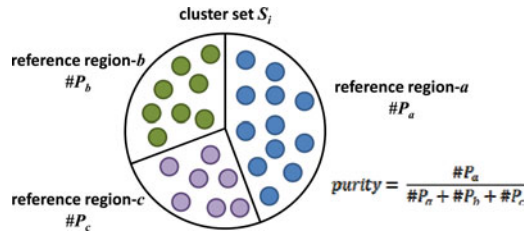
Fig. 6.   Computation of purity value in each cluster. The blue points matched to reference region a are the dominant points in cluster S_i. The purity value of each cluster is the number of dominant points with respect to the total number of points in this cluster.

same color almost randomly distribute on the entire image. The different spatial relationships of matched points can be used to distinguish the positive images containing banknotes from the negative images.

We apply the k-means clustering [12] to analyze the spatial relationship between matched points. The aim of k-means clustering is to partition n matched points into k clusters in which each point belongs to the cluster with the nearest distance. Given a set of matched points $(x_1, x_2, \ldots, x_n)$, where each point is a 2-D position vector, k-means clustering partitions the n points into k sets $(k \leq n)$ $S = \{S_1, S_2, \ldots, S_k\}$ so as to minimize the within cluster sum of distances

$$\arg \min_S \sum_{i=1}^{k} \sum_{x_j \in S_i} \|x_j - \mu_i\|^2 \tag{1}$$

where μ_i is the centroid of S_i, i.e., the mean of points in S_i. In the experiments, we set k as the number of matching regions. For example, if there are four matching regions, then $k = 4$. The results of k-means clustering are presented with *purity* values. Purity is the number of dominant points in a cluster with respect to the total number of points in this cluster. As shown in Fig. 6, in the cluster set S_i, $\#P_a$ that is the number of matched points with reference region a is larger than those from other reference regions, i.e., $\#P_b$ and $\#P_c$. Thus, the points matched to reference region a are dominant points. The purity of this cluster set is the ratio between $\#P_a$ and the sum of $\#P_a$, $\#P_b$, and $\#P_c$. As the matched points of positive images tend to cluster much more than the case in negative images, the purity values of each cluster of positive images are, therefore, much larger than that of negative images. From our empirical observation, if the purity value of any cluster set is less than 0.16, then this query image is determined to be a negative image.

IV. BANKNOTE RECOGNITION SYSTEM DEVELOPMENT

Taking all the components in Fig. 1, we form the component-based banknote recognition system. The system includes three main components: 1) sensors including a camera to capture a query image, a microphone for speech command input, and speakers (or Bluetooth, earphone) for audio output; 2) data capture and analysis to perform command control and banknote recognition by using a wearable computer (e.g., a mini-computer, a personal digital assistant, or a mobile phone); and 3) audio outputs to provide the banknote recognition result.

TABLE II
PROCEDURE OF BANKNOTE RECOGNITION

Components-based Banknote Recognition by SURF

Given a query image, extract SURF descriptors D_q

Matching features thresholding to determine candidate categories
for each banknote category C_i
    for each reference region R_i^j
        $N_i^j :=$ number of matched points between D_i^j and D_q
        if $N_i^j > T_i^j$
            $M_i := M_i + 1$
        endif
    endfor
    if $M_i \geq 2$
        $candidate := candidate \cup C_i$
    endif
endfor

Matching features clustering to filter out false positive recognition
if $candidate = \emptyset$
    $output := no\ bill$
else
    if $|candidate| > 1$
        $candidate := \underset{C_i \in candidtae}{\arg\max} \sum N_i^j$
    endif
    k-means cluster matched points and compute purity value P_i for
    each cluster set S_i
    if $\forall P_i \leq 0.16$
        $output := no\ bill$
    else
        $output := candidate$
    endif
endif

The implementation procedure of our proposed banknote recognition algorithm is summarized in Table II. The SURF is first applied to extract the descriptors D_i^j of the marked reference regions R_i^j (see Fig. 3) in each banknote category C_i, $i = 1, 2, \ldots, 7$. The extracted SURF descriptors are stored in a database for the following matching. Given a query image, the system employs SURF to extract the descriptors D_q, which are then used to match with precomputed SURF descriptors stored in the database. If the number of matched features with reference region j in category i is larger than its associated automatic threshold T_i^j, this reference region is matched to the query image. For each category C_i, if the number of matched regions M_i is not less than 2, C_i is added in the candidate. If no candidate after the matching features thresholding, this means the query image is a negative image. Therefore, the output will show that no bill is detected in the query image. If more than one category is present in the *candidate*, the total number of matched features of all matched reference regions is used to determine the optimal category, which maximizes this total number. Now, only one category is present in the *candidate*.

In order to avoid false positive recognition, we employ k-means clustering to the matched points and obtain a set of clusters and corresponding purity values. Since matched points tend to spread out on the entire negative image, the purity values

of cluster sets in negative cases are very low. Therefore, if any purity value of cluster sets is not large than 0.16, this query image is also determined to be a negative image. Accordingly, the output displays no bill detected. Otherwise, the query image is recognized as a true positive case, and the output returns the banknote value of the query image.

V. Experiments and Analysis

The performance of the proposed banknote recognition algorithm has been evaluated by two datasets where one is collected by developers and the other one is collected by nine blind subjects. This section focuses on the system evaluation by using the dataset collected by developers which includes both positive banknote images and negative background images. The positive set contains a total of 280 images with 40 images covering both front side and back side for each category of bill ($1, $2, $5, $10, $20, $50, and $100). Note that the proposed system recognizes one banknote each time (i.e., there are no multiple banknotes present in the same image.) The negative set contains 100 background images with no banknote in each image. These images are selected from a wide variety of conditions to approximate the environment of real-world application. Experimental results validate the effectiveness and robustness of our component-based banknote recognition algorithm using SURF.

A. Experiment Setup

The banknote dataset is collected from a wide variety of circumstances. Fig. 7 demonstrates four sample images from each condition. The samples from the first row to the sixth row correspond to the conditions of partial occlusion, cluttered background, rotation, scaling, lighting change, and wrinkling. The seventh row displays four sample images of negative images. In some cases, some of the conditions are combined into one condition, such as scaling and cluttering. The dataset will be released to public at our website. In real-world applications, the user may take the banknote in his/her hand which can correspond to the "both occluded and wrinkled" case. The user may also take a negative image with no bill due to the aiming problem. The banknote dataset that is presented in our experiment is more challenging than that from other banknote recognition papers. For example, the dataset in [32] was collected by using a scanner to scan the bills that were taken under restricted or standard conditions. Therefore, our dataset generalizing the conditions of taking banknote images is more challenging and more closely approximates the real-world environment. In the experiments, we randomly select 140 positive images and 50 negative images as training set, which are used to obtain the automatic thresholds for reference regions. The remaining positive and negative images are used as a testing set.

B. Experimental Results and Discussion

Each category of bill images covers all of the conditions of partial occlusion, cluttered background, rotation, scaling change, and illumination change, as well as wrinkling. In the recognition experiments that are evaluated in our testing dataset,
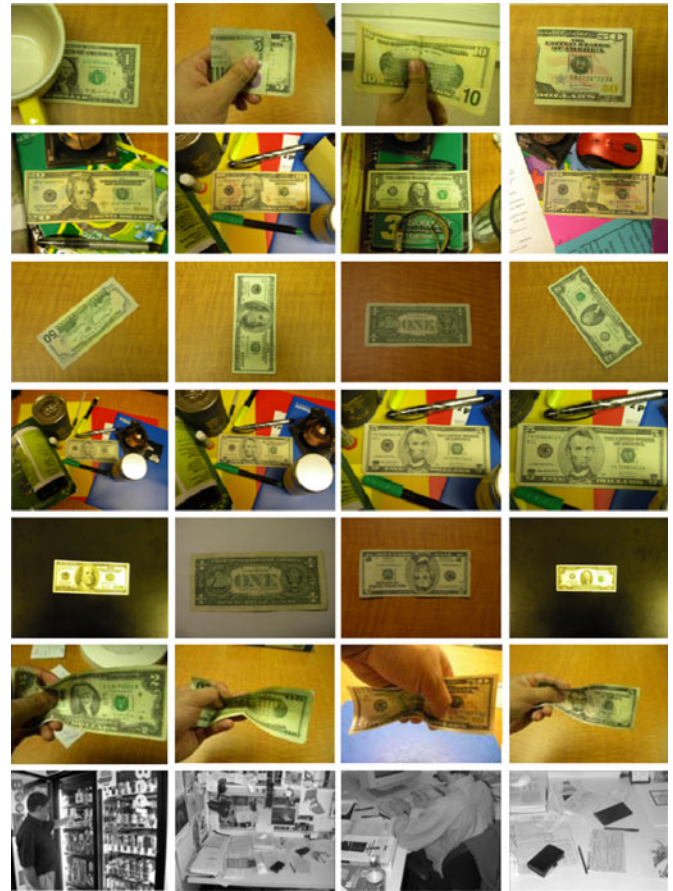

Fig. 7. Training and testing dataset of banknote images taken under different conditions and negative images: first row (partial occlusion), second row (cluttered background), third row (rotation), fourth row (scaling change), fifth row (illumination change), sixth row (wrinkling), and seventh row (negative images).

TABLE III
RECOGNITION RESULTS FOR SEVEN CATEGORIES OF BANKNOTES

Ground Truth	No. of images	False Recognition Rate	True Recognition Rate
$1	20	**0**	**100%**
$2	20	**0**	**100%**
$5	20	**0**	**100%**
$10	20	**0**	**100%**
$20	20	**0**	**100%**
$50	20	**0**	**100%**
$100	20	**0**	**100%**

the proposed algorithm achieves 100% true recognition accuracy for all seven categories and 0% false recognition rate. Table III shows the evaluation results for each category. The experimental results have shown the effectiveness of the monetary features extracted by SURF and our component-based framework for banknote recognition.

Although our algorithm is evaluated in a more challenging dataset, our algorithm achieves much better recognition results and outperforms the existing banknote recognition algorithms.

For instance, the average recognition rate of the algorithm [27] based on SIFT is 76%. Some neural-network-based banknote recognition systems achieved recognition rate no larger than 95% [16], [17], [31]–[33]. Because of specific design, the testing images in these references were captured in highly constrained environments, such as a transaction machine or a specific sensor system which are not portable [5], [31], [33]. Therefore, the variations of testing images evaluated in these systems are quite limited. Their images only presented rotation and translation changes, and sometimes might be worn out or with defected corners. Compared with these references, our system is designed as a portable device for blind users in daily life by using a wearable camera or a camera on a cell phone, and images are captured in open environments with a wide variety conditions including partial occlusion, highly cluttered background, rotation, scaling change, illumination change, as well as wrinkling. In other words, the variety conditions that are presented in our dataset fully cover the ones in others. While our system is evaluated on much more challenging images, our recognition result is still better than other work evaluated on images collected in much simpler conditions.

The success of our proposed algorithm mainly lies in: 1) the scale- and rotation-invariant detector and distinctive descriptor provided by SURF is robust to handle the image rotation, scaling change, and illumination change; 2) our component-based framework generates more category-specific information that is more discriminative to distinguish a query image from cluttered background and effective to deal with partial occlusions; and 3) the spatial clustering approach is effective to filter out negative images by using the spatial relationship between matched points.

However, further experiments show that extremely large scaling change will affect the recognition results. This is because the SURF descriptor is resistant to scaling change in a certain range. When the resolution of a query image is very low, the system cannot detect considerable points in bills. However, in real application, it is reasonable to assume that a blind person takes the bill in the range without extreme scaling change.

The computational cost of the entire recognition algorithm includes extracting SURF features from a query image, matching them with precomputed features of all the reference regions, k-means clustering of matched points, and displaying the recognition output. The average speed of the algorithm on a testing image at the resolution of 1024×768 pixels is about 2.5 s on a computer with 3-GHz CPU.

VI. SYSTEM TESTING BY BLIND USERS

The proposed banknote recognition system was further tested by nine blind subjects including 4 men and 5 women with ages ranging from 36 to 72 years old. The images are captured by a camera mounted on a pair of sunglass with the blind user holding the bill to be recognized as shown in Fig. 8. The dataset collected by blind subjects includes total 579 positive images: 68 images of $1, 44 images of $2, 73 images of $5, 129 images of $10, 71 images of $20, 129 images of $50, and 65 images of $100. As shown in Fig. 9, the images collected by blind subjects

Fig. 8. Evaluation of the proposed system by blind subjects. The bill to be recognized is held by hands and image captured by a camera mounted on a pair of sunglasses.
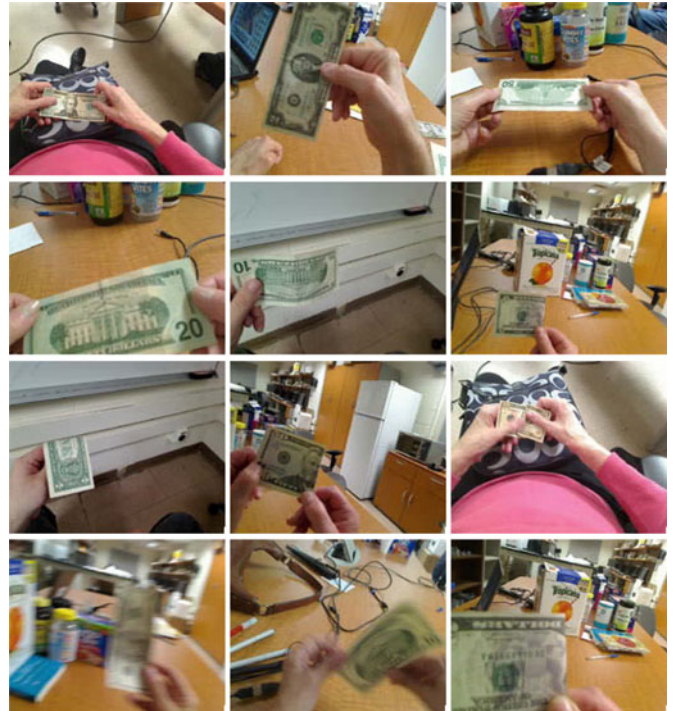

Fig. 9. Sample images of the testing dataset collected by blind subjects. (First row to third row) Banknote images taken by blind user present a variety of conditions including viewpoint change, scaling, rotation, occlusion, wrinkling, and background clutters. (Fourth row) Images with severe motion blur (the first two images) and with high occlusion (the third one).

present a large variation of viewpoint, rotation, occlusion, wrinkling, as well as cluttered background. Our algorithm correctly recognizes all the 579 images and achieves 100% recognition rate.

From the system evaluation by blind subjects, it is observed that the proposed method cannot recognize banknote images in the following two conditions: 1) banknote images taken under severe motion blur (e.g., the first two images of the fourth row in Fig. 9); 2) banknote with only one or none component visible (e.g., the third image of the fourth row in Fig. 9). For highly blurred images, the SURF is not able to extract accurate local image features that are the basis for the flowing feature matching. To balance the reliability of recognition and the ability to handle occlusion, the system requires at least two matched regions, as mentioned in Section III-C. For highly occluded images, they cannot provide enough visible banknote components. However,

our system treats images in both cases as negative recognitions and sends feedback to blind subjects to guide users to adjust their image capture. Since blind users can be collaborative, they will be able to capture good quality images by system feedback and a simple user training process.

VII. Conclusion and Future Work

In this paper, we have proposed a component-based framework for banknote recognition using SURF. Patches with descriptive features for each banknote category are selected as reference regions to match with query images. The proposed algorithm has been evaluated by dataset to a variety of conditions including occlusion, rotation, scaling, cluttered background, illumination change, viewpoint variation, and worn or wrinkled bills, and further by blind subjects. Our approach achieves 100% true recognition rate and 0% negative recognition rate.

In the future work, we will address the issue of motion blur in two perspectives: 1) integrate deblurring techniques [43], [44] into our system; 2) provide a simple training process to familiarize blind users with the device to reduce motion blur. Our future work will also focus on evaluating the proposed system to recognize banknotes of different countries and transferring the method to mobile phones.

Acknowledgment

References

[1] American Foundation for the Blind. (2011). [Online]. Available: http://www.afb.org/

[2] H. Bay, T. Tuytellars, and L. Gool, "SURF: Speeded up robust features," in *Proc. 9th Eur. Conf. Comput. Vis.*, 2006, pp. 404–417.

[3] S. Chae, J. Kim, and S. Pan, "A study on the Korean banknote recognition using RGB and UV information," *Commun. Comput. Inf. Sci.*, vol. 56, pp. 477–484, 2009.

[4] D. Dakopoulos and N. G. Bourbakis, "Wearable obstacle avoidance electronic travel aids for blind: A survey," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 40, no. 1, pp. 25–35, Jan. 2010.

[5] A. Frosini, M. Gori, and P. Priami, "A neural network-based model for paper currency recognition and verification," *IEEE Trans. Neural Netw.*, vol. 7, no. 6, pp. 1482–1490, Nov. 1996.

[6] N. Giudice and G. Legge, "Blind navigation and the role of technology," in *The Engineering Handbook of Smart Technology for Aging, Disability, and Independence*, A. A. Helal, M. Mokhtari, and B. Abdulrazak, Eds. Hoboken, NJ: Wiley, 2008.

[7] C. Harris and M. Stephens, "A combined corner and edge detector," in *Proc. Fourth Alvey Vis. Conf.*, Manchester, U.K, 1988, pp. 147–151.

[8] C. Harris, "Geometry from Visual Motion," in *Active Vision*. Cambridge, MA: MIT Press, 1993.

[9] H. Hassanpour and P. Farahabadi, "Using hidden Markov models for paper currency recognition," *Expert Syst. Appl.*, vol. 36, pp. 10105–10111, 2009.

[10] F. Hasanuzzaman, X. Yang, and Y. Tian, "Robust and effective component-based banknote recognition by SURF features," in *Proc. 20th Annu. Wireless Opt. Commun. Conf.*, 2011, pp. 1–6.

[11] B. Hodges, D. Burton, and M. Uslan, "Show me the money: Access issues: An evaluation of the note teller 2 money identifier. AFB AccessWorld: Technology and People Who Are Blind or Visually Impaired," vol. 8, no. 3, May 2007.

[12] M. Inaba, N. Katoh, and H. Imai, "Applications of weighted Voronoi diagrams and randomization to variance-based K-clustering," in *Proc. ACM 10th Annu. Symp. Comput. Geom.*, 1994.

[13] N. Jahangir and A. Chowdhury, "Bangladeshi banknote recognition by neural network with axis symmetrical masks," in *Proc. IEEE 10th Int. Conf. Comput. Inf. Technol.*, Dec. 2007, pp. 1–5.

[14] Y. Ke and R. Sukthankar, "PCA-SIFT: A more distinctive representation for local image descriptors," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit*, 2004, pp. II-506–II-513.

[15] I. Kocur, R. Parajasegaram, and G. Pokharel, "Global data on visual impairment in the year 2002," *Bull. World Health Org.*, vol. 82, pp. 844–851, 2004.

[16] T. Kosaka and S. Omatu, "Bill money classification by competitive learning," in *IEEE Midnight-Sun Workshop Soft Comput. Methods Ind. Appl.*, 1999, pp. 5–9.

[17] T. Kosaka, S. Omatsu, and T. Fujinaka, "Bill classification by using the LVQ method," in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, 2001, vol. 3, pp. 1430–1435.

[18] J. Lee, S. Jeon, and H. Kim, "Distinctive point extraction and recognition algorithm for various kinds of euro banknotes," *Int. J. Control, Autom., Syst.*, vol. 2, no. 2, pp. 201–206, 2004.

[19] D. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, pp. 91–110, 2004.

[20] C. Lupsa. (2006). Should U.S. bills be blind friendly? The Christian Science Monitor [Online]. Available: http://www.csmonitor.com/2006/1207/p17s02-lign.html

[21] R. Manduchi, J. Coughlan, and V. Ivanchenko, "Search strategies of visually impaired persons using a camera phone wayfinding system," presented at the 11th Int. Conf. Comput. Helping People Spec. Needs, Linz, Austria, 2008.

[22] R. Manduchi, S. Kurniawan, and H. Bagherinia, "Blind guidance using mobile computer vision: A usability study," in *Proc. 12th Int. ACM SIGACCESS Conf. Comput. Accessibility*, 2010, pp. 241–242.

[23] R. Manduchi and S. Kurniawan, "Mobility-related accidents experienced by people with visual impairment," *Insight: Res. Practice Visual Impairment Blindness.*, vol. 4, no. 2, pp. 44–54, 2011.

[24] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 10, pp. 1615–1630, Oct. 2005.

[25] National Research Council, *Electronic Travel Aids: New Directions for Research* (Working Group on Mobility Aids for the Visually Impaired and Blind Committee on Vision), Washington, DC: National Academy Press, 1986, p. 107.

[26] Note Teller 2. (2011). [Online]. Available: http://www.brytech.com

[27] T. Reiff and P. Sincak, "Multi-agent sophisticated system for intelligent technologies," *IEEE 6th Int. Conf. Comput. Cybern.*, Nov. 2008, pp. 37–40.

[28] C. Schimd and R. Mohr, "Local gray value invariants for image retrieval," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 5, pp. 530–535, May 1997.

[29] H. Shen, K. Chan, J. Coughlan, and J. Brabyn, "A mobile phone system to find crosswalks for visually impaired pedestrians," *Technology and Disability*, vol. 20, pp. 217–224, 2008.

[30] H. Shen and J. Coughlan, "Grouping using factor graphs: An approach for finding text with a camera phone," in *Proc. Workshop Graph-based Represent. Pattern Recognit*, 2007, pp. 394–403.

[31] F. Takeda and S. Omatu, "High speed paper currency recognition by neural networks," *IEEE Trans. Neural Netw.*, vol. 6, no. 1, pp. 73–77, Jan. 1995.

[32] F. Takeda, L. Sakoobunthu, and H. Satou, "Thai banknote recognition using neural network and continues learning by DSP unit," in *Proc. 7th Int. Conf. Knowledge-Based Intell. Inf. Eng. Syst.*, 2003, pp. 1169–1177.

[33] F. Takeda and T. Nishikage, "Multiple kinds of paper currency recognition using neural network and application for Euro currency," in *Proc. IEEE-INNS-ENNS Int. Joint Conf. Neural Netw.*, 2000, pp. 143–147.

[34] E. Tekin, J. Coughlan, and H. Shen, "Real-time detection and reading of LED/LCD displays for visually impaired persons," in *Proc. IEEE Workshop Appl. Comput. Vis.*, 2011, pp. 491–496.

[35] Y. Tian, C. Yi, and A. Arditi, "Improving computer vision-based indoor wayfinding for blind persons with context information," in *Proc. 12th Int. Conf. Comput. Helping People Spec. Needs*, 2010, pp. 255–262.

[36] Y. Tian, X. Yang, and A. Arditi, "Computer vision-based door detection for accessibility of unfamiliar environments to blind persons," in *Proc. 12th Int. Conf. Comput. Helping People Spec. Needs*, 2010, pp. 263–270.

[37] A. Vila, N. Ferrer, J. Mantecon, D. Breton, and J. Garcia, "Development of a fast and non-destructive procedure for characterizing and distinguishing

original and fake euro notes," *Analytica Chimica Acta*, vol. 559, pp. 263–273, 2006.

[38] X. Yang, Y. Tian, C. Yi, and A. Arditi, "Context-based indoor object detection as an aid to blind persons accessing unfamiliar environments," *Int. Conf. ACM Multimedia*, 2010, pp. 1087–1090.

[39] S. Yuan, Y. Tian, and A. Arditi, "Clothing matching for visually impaired persons," *Technol. Disability*, vol. 23, pp. 75–85, 2011.

[40] E. Zhang, B. Jiang, J. Duan, and Z. Bian, "Research on paper currency recognition by neural networks," in *Proc. Int. Conf. Mach. Learning Cybern.*, 2003, pp. 2193–2197.

[41] R. Huber, H. Ramoser, K. Mayer, H. Penz, and M. Rubik, "Classification of coins using an eigenspace approach," *Pattern Recognit. Lett.*, vol. 26, pp. 61–75, 2005.

[42] L. Shen, S. Jia, Z. Ji, and W. Chen, "Extracting local texture features for image based coin recognition," *IET Image Process.*, vol. 5, no. 5, pp. 394–401, Aug. 2011.

[43] W. Dong, L. Zhang, G. Shi, and X. Wu, "Image Deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization," *IEEE Trans. Image Process.*, vol. 20, no. 7, pp. 1838–1857, Jul. 2011.

[44] A. Beck and M. Teboulle, "Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems," *IEEE Trans. Image Process.*, vol. 18, no. 11, pp. 2419–2434, Nov. 2009.

[45] S. Shoval, J. Borenstein, and Y. Koren, "Auditory guidance with the Navbelt-a computerized travel aid for the blind," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 28, no. 3, pp. 459–467, Aug. 1998.

[46] A. Huete, J. Victores, S. Martinez, A. Gimenez, and C. Balaguer, "Personal autonomy rehabilitation in home environment by a portable assistive robot," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, no. 99, pp. 1–10, 2011.

[47] N. Aim and J. Arnott, "Computer-assisted conversation for nonvocal people using prestored texts," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 28, no. 3, pp. 318–328, Aug. 1998.

Xiaodong Yang received the B.S. degree from the Huazhong University of Science and Technology, Wuhan, China, in 2009. He is currently working toward the Ph.D. degree in electrical engineering with The City College of New York, New York.

His current research focuses on recognition of objects, texture, and scene categories. His research interests include image processing, object detection, and developing effective image representations for recognition.

YingLi Tian (M'99–SM'01) received the B.S. and M.S. degrees from Tianjin University, Tianjin, China, in 1987 and 1990, respectively, and the Ph.D. degree from the Chinese University of Hong Kong, Hong Kong, in 1996.

After holding a faculty position with the National Laboratory of Pattern Recognition, Chinese Academy of Sciences, Beijing, China, she joined Carnegie Mellon University in 1998, where she was a Postdoctoral Fellow with the Robotics Institute. From 2001 to 2008, she was a Research Staff Member with IBM T. J. Watson Research Center. She is one of the inventors of the IBM Smart Surveillance Solutions. In September 2008, she joined the Department of Electrical Engineering, City College of New York, New York, where she is currently an Associate Professor. Her current research interest focuses on a wide range of computer vision problems from motion detection and analysis to human identification, facial expression analysis, and video surveillance.

Faiz M. Hasanuzzaman is a senior majoring in electrical engineering with The City College of New York (CCNY), New York.

In the summer of 2010, he was a Research for Undergrad Intern with Carnegie Mellon University and the University of Pittsburg, where he was funded by the Quality of life Technology, Human Engineering Research Lab, Pittsburg, PA. In the summer of 2011, he was an Intern with the Rutgers University, NJ, where he was funded by the Department of Homeland Security, Command, Control, and Interoperability Center for Advanced Data Analysis Center. His current research interest focuses on computer vision with the Media Lab, CCNY.