Proceedings of the 5th WSEAS Int. Conf. on SIGNAL, SPEECH and IMAGE PROCESSING, Corfu, Greece, August 17-19, 2005 (pp276-281)

Shape Context for Image Understanding

LEI HE[1], HUAZHOU LIU[2]
[1]Information Technology Department, Armstrong Atlantic State University,
Savannah, GA 31419, US
http://www.computing.armstrong.edu/FacNStaff/leihe/
[1]Electrical & Computer Engineering and Computer Science Department, University of Cincinnati,
Cincinnati OH 45221, US

Abstract: - This paper presents a robust object recognition and recovery method for image understanding using a recent shape feature descriptor: shape context. The novel feature is to unify both object recognition and recovery components into an image understanding system architecture, in which a complementary feedback structure can be incorporated to alleviate processing difficulties of each component alone. The idea is firstly to recognize the preliminary extracted object from a set of models by matching their shape contexts, then to apply the a priori shape information of the identified model for accurate object recovery. The output of the system is the recognized and segmented object. The shape matching method is illustrated by recognizing a set of CAPTCHA and animal silhouette examples with the presence of object translation and scaling, shape deformations and noise. Experiments of object recovery using real biomedical image samples, such as MR knee, have shown satisfactory results.

Key-Words: - Shape Context, Shape Matching, Object Recognition, Object Recovery, Image Understanding

1 Introduction

Image understanding plays an important role in image analysis and computer vision. Generally it includes two key interrelated components: image segmentation and object recognition, as shown in Fig.1. Image segmentation approach such as deformable contour method (DCM) yields contours, either exact or approximate, of objects of interest in images for recognition. Object recognition performs shape matching to identify an object from a set of models. The recognition results can be fed back into the image segmentation to enhance the accuracy and robustness of the segmentation results, which is referred to as object recovery. The objective of this paper is to apply a new shape feature descriptor, shape context [1], for shape matching and object recognition, yielding a robust and efficient object recovery.

The object recognition is implemented by matching the shape contexts of an input DCM [2] contour with those of a set of model contours. The "best" matching contour pair not only determines the correct model class to which the input object belongs, but also constructs the feature point (landmark) correspondences between the input DCM contour and the selected model contour. The correspondences of the contours' segments follow automatically. For the DCM contour segments with a large error compared with the matched model segments, a fine-tuning process, which is formulated as a maximization of a posteriori probability [3], is performed for final object recovery.

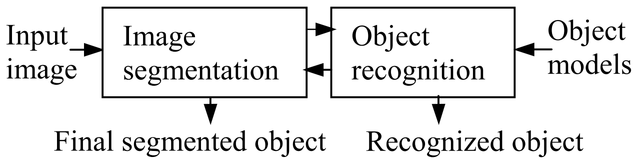


Fig.1 Image understanding process

The rest of the paper is organized as follows. Section 2 briefly reviews DCMs for image segmentation and object recognition methods. The proposed algorithm is presented in Section 3. Experiments on matching animal profiles, recognizing CAPTCHA [17] examples, and recovering shapes in biomedical images are provided in Section 4. Section 5 draws the conclusions.

2 Background

As one of the most advanced and popular image segmentation methods, deformable contour method [2][4-6] iteratively deforms a contour in the image to search the boundary of object of interest. Among variant DCMs, snake method [4] and level set method [2][5][6] are the two most commonly used categories. When the a priori object shape information is available, model-based snakes or level sets can be applied by either embedding the information into the snake energy [7] or level set velocity functions [8] to constrain the admissible contour deformation range. Parameterized deformable models [3][9][10] are commonly used when the a priori object shape information of the object of interest can be represented by a small number of parameters with certain probability distributions. Recent progress in DCMs has advanced the state of the art significantly. However, contour extraction for detailed object recovery still remains a challenge when the contour is not readily present due to noise, blurry contour segments, or complex shapes. Nevertheless, DCMs can in most cases extract the desired object boundary, or a good approximation, for object recognition. Thus our recent work [2] is applied in this application for preliminary image segmentation.

Object recognition is usually implemented by object shape matching methods, such as template matching, statistical classification and structural classification [11]. Template matching is the simplest shape identification method by comparing the input shape to a list of stored shape representations (templates), which is usually formulated as a parameterization problem, with a quadratic fitting criterion to be minimized [12]. It generally can handle only simple cases where there is only a small geometric variation (rotation, size and position variation) between the input shape and templates, thus not suitable to non-rigid shape matching. Statistical approaches [1][13][14] use a set of selected shape measures or features that are more resilient to shape variations to match shapes, such as shape contexts [1] and curvature [13][14], which largely enhances the robustness to object geometric variations and shape deformations. Sclaroff et al. [13] proposed a modal matching method to establish the correspondences of contour feature points and to recognize objects based on the eigenmode description. However, as pointed out in [14], without the connectivity information of the contour, the algorithm is not guaranteed to generate a legal set of correspondences. Hill et al. [14] proposed a three-step algorithm to automatically identify the landmarks on two contours and constructing their correspondences. The main idea is that the ratio of the distance of two contour points with respect to the whole contour arc length should be similar to that of corresponding contour points' distance to the whole arc length of the other contour. The algorithm has good performances on complex object shapes. However, it is computationally demanding due to the iterative processes in all the three steps. Structural classification methods [15][16] match shapes by comparing their structure, i.e., their respective ordered composition of simple sub-patterns or shape primitives. Shapes are represented by such a composition of shape primitives. The selection and segmentation of shape primitives are not easy, and generally depend on the user preferences and experiences. Zhu and Yuille [15] constructed object skeleton model in terms of the principal deformation modes for object recognition. A branch-and-bound approach is applied for skeleton matching based on geometrical features of primitives. Object occlusion and viewpoint variations can be handled by skeleton topology changing operators. The algorithm is sensitive to noise on primitive segmentation and computationally demanding with multiple parameters to be tuned. Shock graph [16] regards object skeleton as a set of singularities (shocks), which can be further represented as a shock tree/graph for shape representation and matching. The matching algorithm is to find the shock graph node correspondences based on both the graphs topological and geometrical similarities, which are constructed by the shock category and attributes, such as location, orientation and time of formation. The shock segmentation and matching algorithms are complex and sensitive to shape noise.

In this paper, we apply a new shape feature descriptor, shape context, to build a robust and efficient image understanding system, which integrates object recognition component using the statistical classification approach and object shape recovery component using the parameterized deformable model.

3 Algorithm Description

As shown in Fig.1, the proposed algorithm includes two steps for object recognition and shape recovery, respectively. Step 1 includes shape context computation and matching for both the input DCM contour and model contours. The output is the identified model class to which the input object belongs to, together with the point correspondences on the matched contours. Based on the matched points, the input and model contours are broken into contour segments. The input contour segments are then matched against the corresponding model segments for error analysis in Step 2. Then for any large error in the segment mismatch, a fine-tuning process utilizing the a priori shape information of the identified model is performed for accurate object shape recovery in the input image.

In Step 1, the input DCM contour and a set of given model contours are firstly sampled to a fixed number of points, e.g., n points from each contour, shape contexts are then computed for the sampled contour points. The shape context is a novel object shape descriptor proposed by Belongie et al. [1], which measures the relative positions between an edge point to other edge points on the object shape. For an edge point p_i, its shape context is computed as a coarse histogram h_i of the relative coordinates of the remaining n-1 edge points:

$$h_i = \# \{q \neq p_i : (q - p_i) \in \text{bin}(k)\}$$

The bins are uniform in log-polar space (k = 1, …, K) [1]. As indicated in [1], the shape context is a robust shape feature descriptor, which is not only invariant to object translation and scaling, but also robust under small geometrical distortions, occlusion and presence of outliers. Thus it was selected for our object recognition and recovery applications. Considering a point p_i on the input DCM contour and a point q_j on a model contour, their shape context cost is formulated as a χ^2 distance measurement:

$$C_{ij} = C(p_i, q_j) = \frac{1}{2}\sum_{k=1}^{K} \frac{[h_i(k) - h_j(k)]^2}{h_i(k) + h_j(k)}$$

where $h_i(k)$ and $h_j(k)$ denote the K-bin normalized histogram at p_i and q_j, respectively. Thus the set of costs C_{ij} between all pairs of points p_i on the input DCM contour and q_j on the model contour becomes a square matrix of matching costs with all possible matches. Our purpose is to select the match that minimizes the total cost of matching,

$$H(\pi) = \sum_i C(p_i, q_{\pi(i)})$$

subject to the constraint that the matching is one-to-one, i.e., π is a permutation. In the matching process, a linear assignment algorithm [18] was applied to minimize the total cost of the matching. The match with the minimum cost value determines both the correct model and the point correspondences between the input contour and the selected model contour. Contour landmarks on both contours are then selected accordingly. At last, the input and model contour segments correspondences can be determined automatically from the contour landmarks correspondences.

Step 2 utilizes the a priori shape information of the identified model to implement the parameterized deformable model for object shape recovery in the input image. The input contour segments are firstly compared with their corresponding model contour segments for shape errors. To compute the shape error E between an input contour segment D_i and its corresponding model contour segment M_j, D_i should be linearly transformed to the coordinate system of M_j as D_i', such as affine transform (translation, rotation and scaling). Then after normalizing D_i' and M_j to be the same length N, the shape error E can be computed as:

$$E(D_i, M_j) = \frac{1}{\mathrm{N}}\sum_{k=1}^{\mathrm{N}} \sqrt{(xD'_{ik} - xM_{jk})^2 + (yD'_{ik} - yM_{jk})^2}$$

, where (xD_{ik}', yD_{ik}') and (xM_{jk}, yM_{jk}) are the coordinates of the kth corresponding points on D_i' and M_j, respectively. The final contour segment correction for the segments with large errors (e.g., E is greater than a threshold) is an application of the procedures in [3] on open curves, which formulates the contour or contour segment searching problem as a maximization the a posteriori probability problem. The elliptical Fourier descriptor is applied to represent the contour segments (open curves), i.e., the parameter vector is like $p=(a_0, c_0, a_1, c_1, \ldots, a_M, c_M)$ with a and c are the elliptical Fourier parameters. More specifically, for the input contour segment $c(x, y)$ corresponding to a model segment $t_p(x, y)$ with p being a parameter vector, we have the following object function to be maximized:

$$\Pr(t_{map} \mid c) = \max_p \Pr(t_p \mid c) = \max_p \frac{\Pr(c \mid t_p)\Pr(t_p)}{\Pr(c)} = M(p)$$

, where t_{map} is the maximum a posteriori solution, $Pr(t_p)$ is the prior probability of t_p, and $Pr(c|t_p)$ is the conditional probability, or likelihood, of the contour segment given the model. The model is obtained by assigning a Gaussian distribution on the mean contour obtained from an atlas or drawn by a domain expert as in [3]. The maximization of the above a posteriori probability function can be simplified to the maximization of the following function with respect to the parameter vector p:

$$M(p) = \sum_{i=1}^{\mathrm{M}} [\ln(\frac{1}{\sigma_i \sqrt{2\pi}}) - \frac{(p_i - m_i)^2}{2\sigma_i^2}] + \frac{k}{\sigma_n^2}\sum_{n=1}^{\mathrm{N}} b(x(p,n), y(p,n))$$

, where m_i and σ_i^2 are the mean and the variance of p_i, σ_n^2 is the noise variance of the input image region including the contour segment; N is the number of points on that segment; $b(x(p, n), y(p, n))$ is an edge map of the input image and k is a constant. The input DCM contour is usually a good approximation of object shape and there is very little variation existing between the DCM contour and the correct object boundary, i.e., the optimal value is close to the initial value. Therefore, the gradient search method can be applied to maximize the a posteriori probability function and the parameter vector p maximizing the function represents the desired object contour segment. The combination of all such contour segments becomes the final segmentation result. In practice, a feedback can be added from Step 2 to Step 1 if the final segmentation result is still not satisfactory, which is generally resulted from a poor DCM input contour due to complex image segmentation problem.

4 Experiments

In this section, the shape context-based shape matching algorithm for object recognition is illustrated using two sets of shapes: the first set includes eight animal silhouettes from [16]; the second set contains a set of CAPTCHA [17] examples. At last, an experiment on two MRI knee images is used to illustrate how to apply the shape matching results for object shape recovery.

In Fig.2, eight biological shapes from [16] are matched with each other to find the most similar shapes to the input shapes. The cross matching results with the normalized cost values are shown in Table 1. It can be seen that the objects belonging to the same category as the input shape are the most similar to it, as shown by the bolded numbers in the table. The matching results show that the shape context is robust to shape deformation and noise, which makes it suitable for non-rigid object recognition.

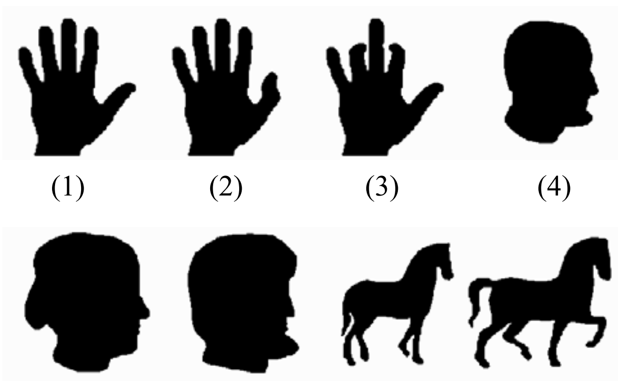


Fig. 2 Biology shapes for shape matching experiments

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(1)	**0**	**.19**	**.239**	.666	.706	.681	.773	.746
(2)	**.19**	**0**	**.238**	.647	.69	.653	.819	.768
(3)	**.239**	**.238**	**0**	.67	.691	.687	.845	.79
(4)	.666	.647	.67	**0**	**.227**	**.27**	.77	.74
(5)	.706	.69	.691	**.227**	**0**	**.29**	.767	.742
(6)	.681	.653	.687	**.27**	**.29**	**0**	.763	.75
(7)	.773	.819	.845	.77	.767	.763	**0**	**.551**
(8)	.746	.768	.79	.742	.742	.74	**.551**	**0**

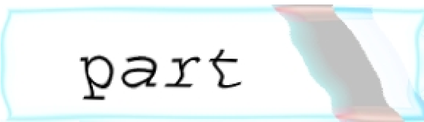
Table 1. Shape matching experiments

In Fig.3, several CAPTCHA (Completely Automated Public Turing Test to Tell Computers and Humans Apart) examples are used to illustrate the application of the shape context-based shape matching approach for character recognition. As indicated in [17], the CAPTCHA was proposed by Manuel Blum's group at Carnegie Mellon University in order to provide practical help to companies like Yahoo! to protect their free email services. The underlying principle behind the design of CAPTCHAs is a reduction to a hard artificial intelligence (AI) problem: human can solve it easily, but computer programs cannot. If the problem cannot be solved by computer programs, it can be used as a CAPTCHA to improve system security and screen out "bots[1]". If it can be solved, then it marks the scientific progress on a hard AI problem. In this experiment, we apply the shape context-based shape matching method to break one category of CAPTCHA test: EZ-Gimpy, with different types of examples shown in Fig.3.

Given an EZ-Gimpy example, a threshold method is firstly applied to detect the edge points in the image. Then the location of the word to be recognized in the image can be determined according to the fact that the region containing word has much more edge points than other regions. A small scanning window (around half size of a letter) is applied to search the candidate sub-regions over the image. The sub-regions containing more edge points than others and connected with similar neighbors are merged together and the resultant big region is considered as the word region. Thirdly, the word is divided into individual letters with the assumption that the font size is fixed in all the images, i.e., each letter occupies the same space in images. After character separation, character recognition is implemented by matching image characters with model templates based on their shape contexts. In this step, each character in an image may has several possible candidates, which are saved in a list according to their matching costs. Thus there are multiple possible words in the image. At last, the final word is determined according to the recognized letters by searching the best match from a given dictionary. The recognition results of different types of EZ-Gimpy with cost values are shown in Fig. 3. This experiment further illustrates the shape context robustness to object shape noise and deformations in object recognition applications.

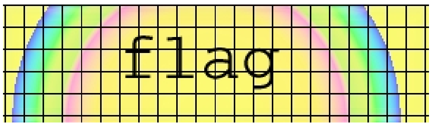
[1] A software program that imitates the behavior of a human, as by signing up thousands of email accounts every minute.

(a) **much: 26.47**
such: 30.53
moon: 40.08
sock: 42.56



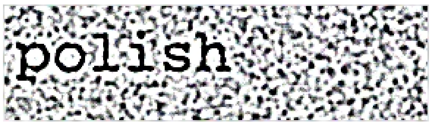
(b) **part: 32.98**
cart: 37.63
past: 42.33
sort: 46.3

(c) **bell: 21.64**
ball: 28.3
tall: 36.35
tail: 40.74



(d) **flag: 22.62**
flat: 35.55
ring: 41.78
free: 51.59

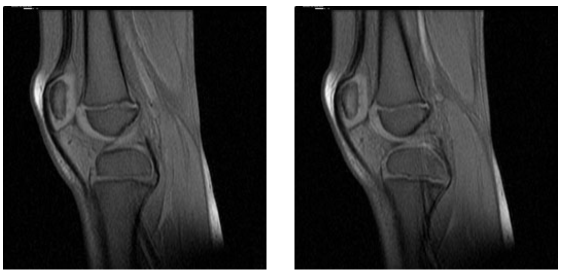
(e) **past: 49.53**
east: 50.37
test: 52.28
rest: 53.3



(f) **polish: 75.02**
poison: 80.79

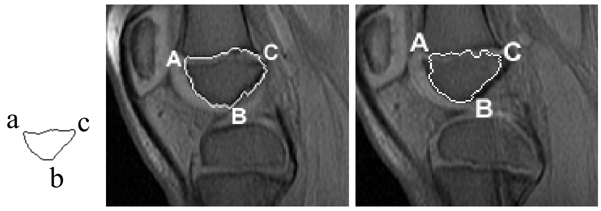
Fig. 3 Shape context for breaking CAPTCHA examples

For object recovery, an experiment on two MRI knee images is used to illustrate the contour segment correction algorithm. The images in Fig.4(a) and 4(b) are two examples of midline sagittal MRI knee images of size 256 by 256. To extract the femoral condyle (top portion of the knee), the goal is to delineate the top segment of the contour that separates the semicircular portion of the femur from the stem. The challenge is that there is a blurry edge segment along the middle top boundary, while the left and right portions of the femoral condyle are rather darker than the middle region. This prevents the deformable contour to reach the real boundary on the two sides before it flows out from the top.



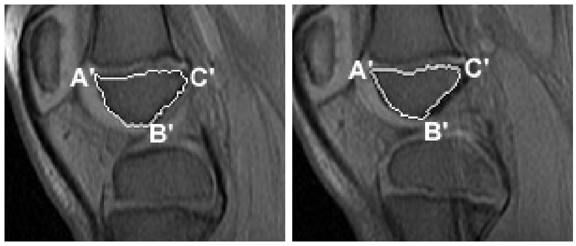
(a) Knee 1 (b) Knee 2

Fig.4 Input MRI knee images



(a) Model (b) Knee 1 input (c) Knee 2 input

Fig.5 The knee model and input shapes obtained by a DCM [2]



(a) Knee 1 final result (b) Knee 2 final result

Fig.6 The knee recovery results after contour segments correction

Before Correction		After Correction	
Figure 5(b)		Figure 6(a)	
Segment Error	Segment Points #	Segment Error	Segment Points #
0.624	42 (AB)	0.333	39 (A'B')
4.217	40 (BC)	1.629	32 (B'C')
2.458	48 (CA)	1.109	52 (C'A')
Figure 5(c)		Figure 6(b)	
Segment Error	Segment Points #	Segment Error	Segment Points #
2.435	43 (AB)	1.18	39 (A'B')
1.483	31 (BC)	1.073	31 (B'C')
2.115	45 (CA)	1.267	47 (C'A')

Table 2. The contour segments errors comparison before and after final correction

As stated in the Step 1 of Section 3, two input knee contours (Fig.5(b), (c)) obtained by the DCM [2] are firstly matched with the knee model in Fig.5(a) to construct the correspondences of contour points. Here three pairs of landmarks are selected on the input and model contours, as shown in Fig.5. The model landmarks *a*, *b*, and *c* correspond to the A, B and C landmarks on the input contours. The contour segment correspondences follow automatically. For the first input contour in Fig.5(b), two large error segments of BC and CA, which indicate the segmentation difficulties mentioned earlier, have to be further refined as described in the Step 2 of Section 3. The constructed final result is shown in Fig.6(a) with both segments BC and CA being corrected and shown as segments B'C' and

C'A'. Similarly, for the second input contour in Fig.5(c), two large error segments of AB and CA are corrected and shown as segments A'B' and C'A' in Fig.6(b). The contour segment errors before and after correction are listed in Table 2. It can be seen the final results have less shape error than those before correction.

5 Conclusion

In this paper, a robust and efficient shape context-based shape matching method is presented to solve the object recognition and recovery problems for image understanding. The presented image understanding system incorporates a complementary feedback structure to alleviate processing difficulties in each of the object recognition and shape recovery components alone. The object recognition is implemented by a statistical classification approach, in which a shape context-based shape matching method is applied to identify the preliminary extracted object from a set of models. It also constructs the contour landmark correspondences on the input and model contours. The correspondences of the contours' segments follow automatically. The identified model shape information can be utilized for the object recovery as the a priori knowledge, which is realized by a parameterized deformable model. For the input contour segments with a large error when compared with their corresponding model segments, a fine-tuning process, which is formulated as a maximization of a posteriori probability, is performed for the segments correction. The output of the system is the recognized and segmented object in the input image. The experiments with the animal shape matching and CAPTCHA recognition and the MRI knee shape recovery demonstrate the capability and potential of this system.

References:

[1] S. Belongie, J. Puzhicha, and J. Malik, "Shape matching and object recognition using shape contexts", *IEEE Trans. on PAMI,* Vol. 24, No. 4, 2002, pp. 509-522.

[2] X. Wang, L. He, and W.G. Wee, "Deformable contour method: a constrained optimization approach", *Int. J. Comput. Vision,* Vol. 59, No. 1, 2004, pp. 87-108.

[3] L.H. Staib and J.S. Duncan, "Boundary finding with parametrically deformable models", *IEEE Trans. on PAMI,* Vol. 14, No. 11, 1992, pp. 1061-1075.

[4] M. Kass, A. Witkin, D. Terzopoulos, "Snakes: active contour models", *Int. J. Comput. Vision,* Vol. 1, No. 4, 1998, pp. 321-331.

[5] R. Malladi, J. Sethian, B. Vemuri, "Shape modeling with front propagation: a level set approach", *IEEE Trans. on PAMI,* Vol. 17, No. 2, 1995, pp. 158-171.

[6] V. Caselles, R. Kimmel, G. Sapiro, "Geodesic active contours", *Int. J. Comput. Vision,* Vol. 22, No. 1, 1997, pp. 61-79.

[7] B. Ostlad and A. Tonp, "Encoding of a priori information in active contour models", *IEEE Trans. on PAMI,* Vol. 18, 1996, pp. 863-872.

[8] M. Rousson and N. Paragios, "Shape priors for level set representations", *Proc. Seventh ECCV*, 2002, pp. 78-92.

[9] A.K. Jain, Y. Zhong, and S. Lakshmanan, "Object matching using deformable templates", *IEEE Trans. on PAMI,* Vol. 18, No. 3, 1996, pp. 267-278.

[10] A. Chakraborty and J.S. Duncan, "Game-theoretic integration for image segmentation", *IEEE Trans. on PAMI*, Vol. 21, No. 1, 1999, pp. 12-30.

[11] R.C. Gonzalez and R.E. Woods, *Digital image processing*, 2nd Ed., Prentice Hall, 2002.

[12] I. Cohen, N. Ayache and P. Sulger, "Tracking points on deformable curves," *Proc. Second ECCV,* 1992, pp. 458-466.

[13] S. Sclaroff and A.P. Pentland, "Modal matching for correspondence and recognition," *IEEE Trans. on PAMI,* Vol. 17, No. 6, pp. 545-561, Jun. 1995.

[14] A. Hill, C.J. Taylor, and A.D. Brett, "A framework for automatic landmark identification using a new method of nonrigid correspondence," *IEEE Trans. on PAMI,* Vol. 22, No. 3, pp. 241-251, Mar. 2000.

[15] S. Zhu and A.L. Yuille, "FORMS: a flexible object recognition and modeling system", *Int. J. Comput. Vision*, Vol. 20, 1996, pp. 187-212.

[16] K. Siddiqi, A. Shokoufandeh, S.J. Dickinson, and S.W. Zucker, "Shock graphs and shape matching," *Int. J. Comput. Vision,* Vol. 35, No. 1, 1999, pp. 13-32.

[17] G. Mori and J. Malik, "Recognizing objects in adversarial clutter: breaking a visual CAPTCHA," *IEEE Int. Conf. on CVPR*, June 2003, pp. 134-141.

[18] R. Jonker and A. Volgenant, "A shortest augmenting path algorithm for dense and sparse linear assignment problems," *Computing*, Vol. 38, 1987, pp. 325-340.