Sparse Representations for Efficient Shape Matching

Alexandre Noma
Department of Computer Science
IME-USP, University of São Paulo
São Paulo, Brazil
Web page: www.vision.ime.usp.br/∼noma

Roberto M. Cesar-Jr
Department of Computer Science
IME-USP, University of São Paulo
São Paulo, Brazil
Web page: www.ime.usp.br/∼cesar

Abstract—**Graph matching is a fundamental problem with many applications in computer vision. Patterns are represented by graphs and pattern recognition corresponds to finding a correspondence between vertices from different graphs. In many cases, the problem can be formulated as a quadratic assignment problem, where the cost function consists of two components: a linear term representing the vertex compatibility and a quadratic term encoding the edge compatibility. The quadratic assignment problem is NP-hard and the present paper extends the approximation technique based on graph matching and efficient belief propagation, described in [1], by using sparse representations for efficient shape matching. Successful results of recognition of 3D objects and handwritten digits are illustrated, using COIL and MNIST datasets, respectively.**

Keywords-**point pattern matching; graph matching; quadratic assignment; Markov random fields; efficient belief propagation; sparse shape representations; shape metric; 3D object recognition; handwritten digits;**

I. INTRODUCTION

Graphs are extensively used to represent complex structures. Many problems can be formulated as an attributed graph matching problem: vertices correspond to local features of the image and edges correspond to relations between features. Both vertices and edges can be attributed, encoding feature vectors. Graph matching consists of finding a correspondence between vertices from two given graphs, here denoted as model and input. The model graphs represent the classes and the input graphs represent the patterns to be classified. The key idea is to compute a mapping between input and model vertices, minimizing a global dissimilarity measure between the attributes from the two graphs.

In many cases, the problem can be formulated as a quadratic assignment problem and we want to find a correspondence which minimizes an energy consisting of a linear term to evaluate vertex attributes, and a quadratic term to evaluate edge attributes. Here, the linear term evaluates features representing the 'appearance' (e.g. shape contexts), while the quadratic term evaluates the 'structure' (e.g. spatial relations).

Recently, learning schemes (e.g. [2], [3]) have been proposed to improve the classification, requiring efficient approaches for shape matching (SM). In the present paper, based on efficient belief propagation (BP) [4], we explore the

spatial relations between points sampled from the contours of an object (e.g. computed using the Canny edge detector) to obtain compact representations for efficient SM.

For graphs with loops, there has been little theoretical understanding of the max-product BP algorithm, as observed in [5], [6]. In the colorization experiments of [1], the graph edges were created between adjacent regions, producing a considerable number of loops, in which messages may circulate indefinitely and the BP algorithm may not converge to a stable equilibrium [7]. As a result, the colorizations computed in [1] depended on fine tunings of the parameter λ_1 in the energy function. Although we follow the same framework described in [1], the compact representations proposed in this work correspond to trees or single loop graphs, for which the max-product BP is known to converge to a stable fixed point or a periodic oscillation, respectively, in which the BP algorithm is known to have good performance [5], [6].

Closely related to our work are those due to Belongie et al. [8] and Torresani et al. [9]. As in [8], we use shape contexts for the appearance. For the structure, our compact representations explore the spatial relations through adjacency, distance and orientation between vertices. The authors of [9] also used these three aspects of spatial relations, but in a complex energy formulation, optimized by a dual decomposition technique based on exhaustive search for local subproblems. As in [1], the present paper explores spatial relations using a simple (yet very useful) quadratic assignment formulation based on Markov random fields (MRFs).

The main contributions of this work are: a new MRF-based quadratic assignment approach for SM, proposing new sparse representations for shapes through a Markov component, together with a new metric for shape distance, based on the computed beliefs. The core of the proposed method relies on the efficient graph matching based on BP, described in [1], producing successful results. Experiments are illustrated using two well-known datasets, COIL [10] and MNIST [11], for recognition of 3D objects and handwritten digits, respectively. Differently from [1], for all experiments, we used the same value for the parameter λ_1 ($= 0.6$).
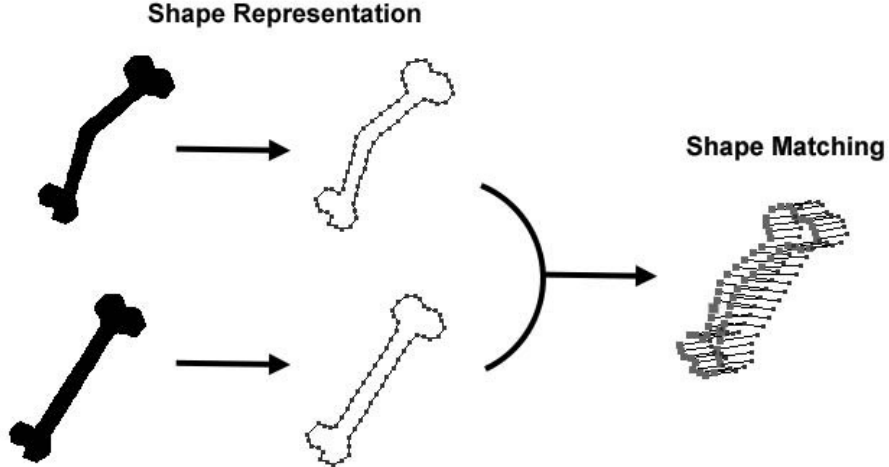
This paper is organized as follows. In Section II, we

Figure 1. Overview of our shape matching approach.

formulate the generic graph matching problem as MRFs. Section III describes the optimization based on BP. Section IV is dedicated to the sparse representation for shapes, the proposed shape distance and the specification of the energy terms. Experimental results are illustrated in Section V. Finally, some conclusions are drawn in Section VI.

II. GRAPH MATCHING AS MRFS

Let $G = (V, E, \mu, \nu)$ be an attributed graph. V is the set of vertices. $E \subseteq V \times V$ is the set of edges. μ assigns an attribute vector to each vertex of V. Similarly, ν assigns an attribute vector to each edge of E. We focus on matching two graphs, an input graph G_i representing a pattern to be classified, and a model graph G_m representing a prototype associated to a class. Given $G_i = (V_i, E_i, \mu_i, \nu_i)$ and $G_m = (V_m, E_m, \mu_m, \nu_m)$, we define a MRF on G_i. For each input vertex $p \in V_i$, we want to associate a model vertex $\alpha \in V_m$. The quality of a labeling $f : V_i \rightarrow V_m$ is given by:

$$E(f) = \sum_{p \in V_i} D_p(f_p) + \lambda_1 \sum_{(p,q) \in E_i} M(f_p, f_q) , \quad (1)$$

where λ_1 weighs the influence of the quadratic term. Each vertex has an attribute vector $\mu_i(p)$ in G_i and $\mu_m(\alpha)$ in G_m. Let $f_p \in V_m$ be the label of $p \in V_i$. The linear term $D_p(f_p)$ compares $\mu_i(p)$ with $\mu_m(f_p)$, assigning a cost proportional to the vertices dissimilarity. Each (directed) edge in each graph has an attribute vector $\nu_i(p, q)$ in G_i and $\nu_m(\alpha, \beta)$ in G_m, where $(p, q) \in E_i$ and $(\alpha, \beta) \in E_m$. The Markov component $M(f_p, f_q)$ compares $\nu_i(p, q)$ and $\nu_m(f_p, f_q)$, assigning a cost proportional to the edges dissimilarity.

An important work involving MRF and graph matching is due to Caelli and Caetano [12], in which the MRF is defined on the model graph G_m, assuming that there is one 'signal' (model) embedded in the 'scene' (input, consisting of the signal plus a set of noisy vertices). The method

was applied to the subgraph isomorphism of straight line segments. In [1], the authors propose an important extension to generalize the method capabilities, defining the MRF on the input graph G_i: each input vertex is labeled and the solution can be generalized for both homomorphism (many-to-one) and maximum common subgraph (one-to-one). Anguelov et al. [13] also formulate the correspondence problem as a 'mapping from the scene to the model', in which case, they inverted the roles assuming that the scene is a 'partial or a complete view' of the model. In contrast, as in [1], we assume that the model is a partial or full view of the scene.

III. OPTIMIZATION BASED ON BP

Finding a labeling that minimizes Equation 1 corresponds to the maximum a posteriori (MAP) estimation problem. We use the max-product BP for optimization, which is formulated via probability distributions. An equivalent computation is performed using negative log probabilities, where the max-product becomes a min-sum, which is less sensitive to numerical artifacts and directly corresponds to the energy in Equation 1.

The method works by passing messages around the graph according to the connectivity given by the edges. Each message is a vector with dimension given by the number of possible labels $|V_m|$. Let m_{pq}^t be the message that vertex p sends to a neighbor q at iteration t. Initially, all entries in m_{pq}^0 are zero and at each iteration new messages are computed by:

$$m_{pq}^t(f_q) = \min_{f_p} \left(M(f_p, f_q) + D_p(f_p) + \sum_{s \in E_i(p) \backslash \{q\}} m_{sp}^{t-1}(f_p) \right) \quad (2)$$

where $E_i(p) \setminus \{q\}$ denotes the neighbors of p except q. After T iterations, for each input vertex, a belief vector is computed representing the costs for each possible label:

$$b_q(f_q) = D_q(f_q) + \sum_{p \in E_i(q)} m_{pq}^T(f_q) \ . \qquad (3)$$

For each input vertex, by choosing a label with minimum cost we obtain a homomorphism. In order to satisfy the two-way constraints, a simple post-processing was applied: for each model vertex, only the cheapest input vertex was kept in the solution, associating to a NULL label the remainder to indicate they are not classified.

A. Efficient computation via min-convolution

Felzenszwalb and Huttenlocher [4] proposed several techniques to improve the running time of BP. They noticed that each message update can be expressed as a min-convolution, which was applied on quadratic terms representing smoothness constraints for low level vision problems. In [1], the authors extended this technique to explore spatial relations in terms of adjacency, distance and orientation between points for point matching.

Firstly, Equation 2 can be rewritten as a min-convolution:

$$m_{pq}^t(f_q) = \min_{f_p} \left(M(f_p, f_q) + h(f_p) \right) \ , \qquad (4)$$

where $h(f_p) = D_p(f_p) + \sum m_{sp}^{t-1}(f_p)$. This is analogous to the standard discrete convolution operation, replacing the sum by a product and the min operator by a sum. As observed in [4], while standard discrete convolutions can be efficiently computed using the FFT, no such general result is known for min-convolutions. However, for commonly used smoothness constraints, the authors showed how to compute the BP messages updates in linear time. Following this idea, in [1], the authors efficiently computed the messages by assuming:

$$m_{pq}^t(f_q) = \min \left(H(f_q), \min_{f_p} h(f_p) + d \right) \ . \qquad (5)$$

which was based on the Potts model described in [4]. However, there is a fundamental difference relied on $H(f_q)$:

$$H(f_q) = \min_{f_p \in E_m(f_q) \cup \{f_q\}} \left(h(f_p) + M(f_p, f_q) \right) \ , \qquad (6)$$

where the Markov component $M(f_p, f_q)$ was used to compare input and model edges in terms of lengths and orientations, and to penalize the non-preserving adjacencies (in our case, the adjacencies between points in a sparse representation of the shape). Note that the amortized computational complexity for each message vector update can be upper bounded by the number of edges in a sparse model graph, leading to an efficient computation for the message updates.

IV. SHAPE MATCHING (SM)

Shape is a significant cue for queries into pictorial databases and a great deal of research on shape similarity has been done using silhouettes (e.g. [14]). More general strategies represent shapes as sets of points which also consider the 'internal' contours of the object. We followed the second approach, where an object is a set of points and its shape is represented by a discrete set of points sampled from the contours of the object computed by an edge detector (e.g. Canny). The goal of SM is to find a correspondence between points from two given shapes, using compact representations.

A. Sparse representation for shapes

For each graph, we roughly sampled uniform spaced points from the contours of an object, representing each point by a vertex. Similar to [15], polygonal contours are used to approximate the actual contours of an object, as shown in Figure 2(c). However, the author of [15] used a triangulation of the points, sampled from the silhouette of the object, to provide a sophisticated decomposition of the object into parts. Here, we use a simpler representation, although not restricted to simple closed curves.

B. Shape distance

In order to operationalize the notion of shape similarity, we propose the following metric:

$$\text{dist}(G_i, G_m) = \sum_{p \in V_i : f_p \neq \text{NULL}} b_p(f_p) + \sum_{p \in V_i : f_p = \text{NULL}} \Lambda_p(f_p) \ , \qquad (7)$$

where the first term is the sum of the computed beliefs (given by Equation 3) considering only the labeled input vertices, and the second term penalizes the vertices without correspondence, assigning the maximum cost $\Lambda_p(f_p) = \max_{p \in V_i : f_p \neq \text{NULL}} \{b_p(f_p)\}$.

Before proceeding to the experiments, we must define each term of Equation 1 for the SM problem.

C. Linear term

Belongie et al. [8] proposed the shape contexts (SCs) as rich descriptors for the appearance information to reduce the ambiguity in the classification. Basically, the SC provides a semi-global description of the spatial distribution of neighboring points by counting the number of points in radial regions yielding histograms, in which each bin represents a different region. For each vertex v, $\mu(v) = \text{SC}(v)$, where $\text{SC}(v)$ is the SC computed on the point represented by v. As in [8], our linear term evaluates the SCs using the χ^2 distance:

$$d_{\chi^2}(\text{SC}(u), \text{SC}(v)) = \frac{1}{2} \sum_k \frac{[h_u(k) - h_v(k)]^2}{h_u(k) + h_v(k)} \ , \qquad (8)$$

where u and v are vertices representing points, and $h_u(k)$ and $h_v(k)$ correspond to the probability of the k-th bin in each histogram associated to $\text{SC}(u)$ and $\text{SC}(v)$.
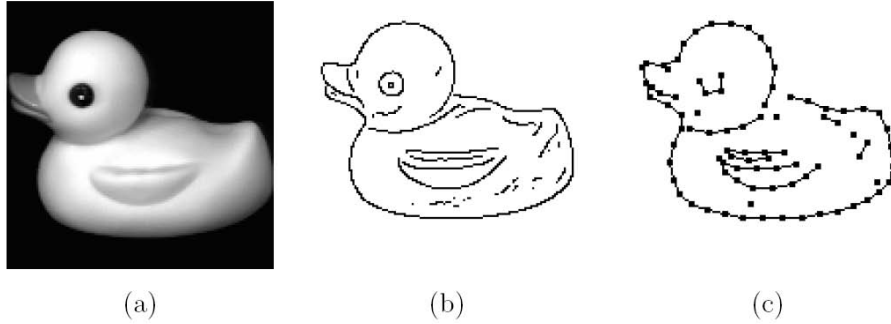
Figure 2. (a) Input object. (b) Computed contours using Canny edge detector. (c) Sparse graphical representation for (b).

D. Quadratic term

For each (directed) edge $e \in E$, a single edge attribute $\nu(e)$ is defined as the (normalized) vector corresponding to the edge. Using the same geometric penalty functions defined in [1], the Markov component compares edges attributes, evaluating pairs of vectors in terms of angle and lengths in order to characterize the spatial relations:

$$c_E(\vec{v_1}, \vec{v_2}) = \lambda_2 \frac{|\cos\theta - 1|}{2} + (1 - \lambda_2)\big||\vec{v_1}| - |\vec{v_2}|\big| , \quad (9)$$

where θ is the angle between the two vectors $\vec{v_1}$ and $\vec{v_2}$, $|.|$ denotes the absolute value, $|\vec{v}|$ denotes the length of $\vec{v}$ (assuming all lengths $|\vec{v}|$ are normalized between 0 and 1), and λ_2 is a parameter to weight the importance between the two terms. The Markov component $M(f_p, f_q)$ is defined as:

$$M(f_p, f_q) = \begin{cases} c_E\big(\nu_i(p,q), \nu_m(f_p, f_q)\big), & \text{if } (f_p, f_q) \in E_m \\ d, & \text{if } (f_p, f_q) \notin E_m \text{ and } f_p \neq f_q \end{cases}$$

$$(10)$$

where the first case compares the respective vectors using Equation 9, and the second penalizes the energy with a positive constant d, encouraging adjacent vertices to have the same label. For the experiments, we used $d = 1$, corresponding to the maximum geometric penalty given by Equation 9.

V. EXPERIMENTS

Similar to [8], our method falls into the category of prototype-based recognition, where classes are represented by ideal examples rather than a set of formal logical rules. The prototype-based recognition can be readily translated into the computational framework of nearest-neighbor methods using multiple stored views. In this case, for 1-NN classifiers, it is important to study their performance for different values of n (specially for small n) where n is the number of prototypes (examples in the training set). For the K-NN classifiers with a fixed training set, it is interesting to analyze their robustness according to different values of K. These two issues are addressed in the present paper. Our experiments were divided into two parts. Firstly, using the COIL dataset for 3D object recognition, we tested our 1-NN
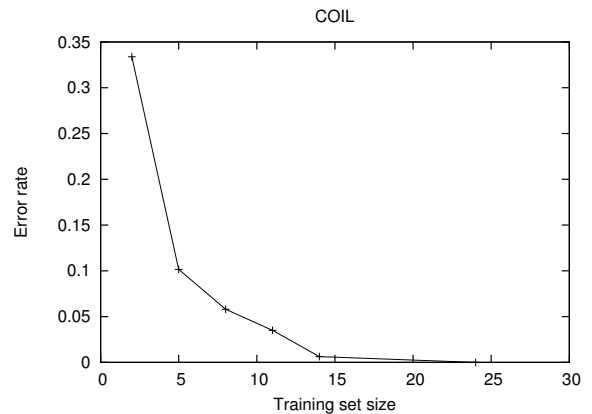

Figure 3. 3D object recognition error rates using 1-NN with different training set sizes.

classifier using sets of different sizes with equally spaced views as prototypes. Then, using the MNIST dataset for handwritten digits recognition, we tested the proposed K-NN approach using different values for K.

A. COIL

This database [10] involves 20 common household objects. Each object was placed on a turnable and photographed every $5°$. This dataset includes 70 different views per object. We tested our 1-NN classifier using training sets with equally spaced views. Figure 3 shows our results: as the number of prototypes increases, the error rate decreases. For instance, using 8 equally spaced views per object, the error rate was 0.058. In a prototype-based approach, different classes/categories need different numbers of views, which depends on the complexity of the given object. Using a K-medoids clustering strategy, we followed a similar approach to the one described in [8] to choose more suitable prototypes. In this case, we improved the error rate to 0.0161 (20 errors from a total of 1242 classifications), using an average of 8 prototypes per object, which is smaller than the error rate in [8] (0.024 using an average of 4 prototypes
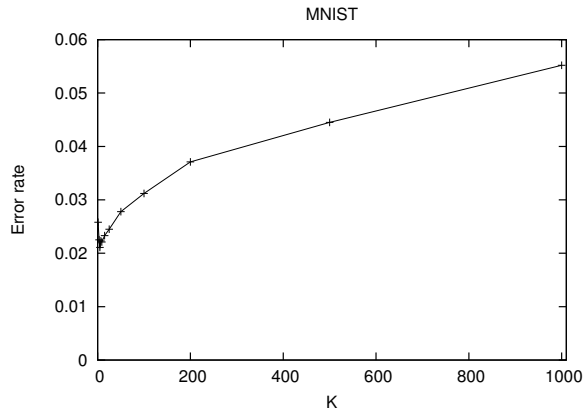
Figure 4. Handwritten recognition error rates using K-NN with different values of K.

per object). Figure 5 illustrates some incorrect classifications produced by our approach, where the models have too many points, in which our shape metric (Equation 7) is not appropriate because it does not penalize the unused labels, assuming that the model is a partial or full view of the scene.

B. MNIST

This dataset [11] consists of 60,000 training and 10,000 test handwritten digits. In http://yann.lecun.com/exdb/mnist/, there is a comparison between more than 60 algorithms, with error rates ranging from 0.0039 to 0.12. Our error rate using 5-NN with all the 60,000 training examples is 0.0211 (211 errors from a total of 10,000 classifications, performing 600,000,000 graph matchings), which is smaller than K-NN with L_2 distance (0.0309), demonstrating the importance of the structural information. Figure 4 illustrates the behaviour of our K-NN classifier for different values of K, where we achieved reasonable error rates even for high values of K, which suggests that our matching process was very robust. Figure 6 illustrates some of the errors produced by our approach. Some of them are very ambiguous, e.g. digit '6' similar to '0', in the first row. Others, like '2' and '8', '3' and '8', '9 and '8', suffer from the same restriction pointed out in the COIL experiment, where Equation 7 does not penalize the unused labels.

VI. CONCLUSIONS

We have presented a new shape representation, which is sparser than the representation based on inner-distances [16] used in [2]. A key characteristic of our method is the fact that it explores the spatial relations between points sampled from the contours of an object in order to represent a polygonal approximation of the shape. This approach leads to an efficient, simple and useful quadratic assignment formulation, where the spatial relations are represented as a Markov component in a MAP-MRF framework. The proposed method obtained successful results in recognizing

3D objects and handwritten digits. In order to improve the classification, it would be interesting to investigate the possiblity of combining the proposed approach with a learning scheme (e.g. [2], [3]).

REFERENCES

[1] A. Noma, L. Velho, and R. M. Cesar-Jr, "A computer-assisted colorization approach based on efficient belief propagation and graph matching," in *CIARP '09: Proceedings of the 14th Iberoamerican Conference on Pattern Recognition*. Berlin, Heidelberg: Springer-Verlag, 2009, pp. 345–352.

[2] L. Chen, J. J. McAuley., R. S. Feris, T. S. Caetano, and M. Turk, "Shape classification through structured learning of matching measures," in *IEEE Conf. on Computer Vision and Pattern Recognition*, June 2009, pp. 365–372.

[3] T. S. Caetano, J. J. McAuley, L. Cheng, Q. V. Le, and A. J. Smola, "Learning graph matching," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 6, pp. 1048–1058, 2009.

[4] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient belief propagation for early vision," *Intl. J. Comput. Vision*, vol. 70, no. 1, pp. 41–54, 2006.

[5] Y. Weiss, "Correctness of local probability propagation in graphical models with loops," *Neural Comput.*, vol. 12, no. 1, pp. 1–41, 2000.

[6] Y. Weiss and W. T. Freeman, "On the optimality of solutions of the max-product belief-propagation algorithm in arbitrary graphs," *IEEE Transactions on Information Theory*, vol. 47, no. 2, pp. 736–744, February 2001.

[7] J. Pearl, *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann, 1988.

[8] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 4, pp. 509–522, April 2002.

[9] L. Torresani, V. Kolmogorov, and C. Rother, "Feature correspondence via graph matching: models and global optimization," in *Proc. of the European Conference on Computer Vision*. Berlin, Heidelberg: Springer-Verlag, 2008, pp. 596–609.

[10] H. Murase and S. K. Nayar, "Visual learning and recognition of 3-d objects from appearance," *Intl. J. Comput. Vision*, vol. 14, no. 1, pp. 5–24, 1995.

[11] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," in *Proc. of the IEEE*, 1998, pp. 2278–2324.
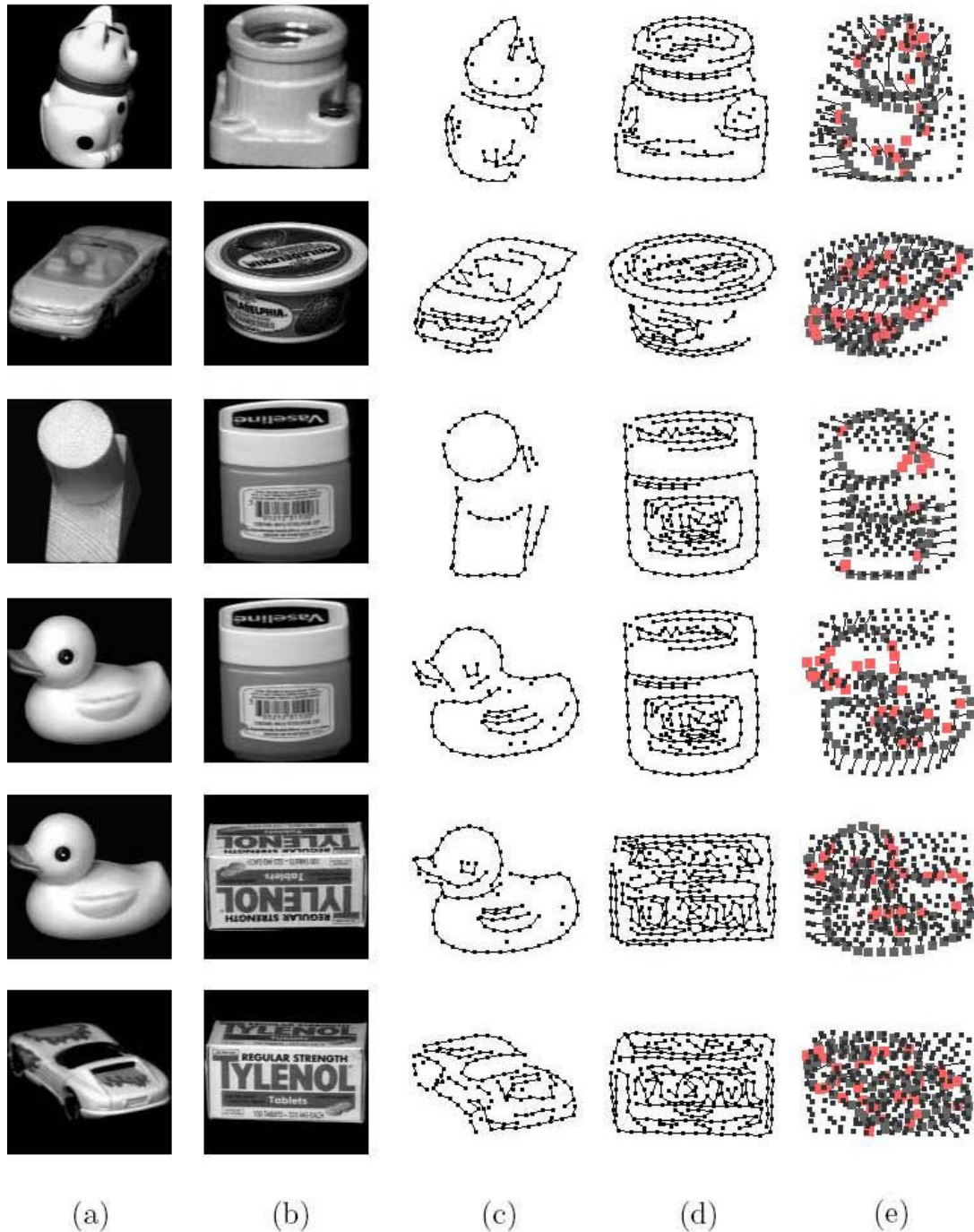
Figure 5. Some incorrect classifications using the COIL dataset by our 1-NN approach, in which the models have too many points, without any chance to be a partial or full view of the scenes. (a) Input image. (b) Nearest neighbor (model), according to the shape metric defined in Equation 7. (c) Sparse graph representation for (a). (d) Similarly, sparse representation for (b). (e) Shape matching, where the input vertices without correspondence are highlighted.

[12] T. Caelli and T. Caetano, "Graphical models for graph matching: approximate models and optimal algorithms," *Pattern Recognition Letters*, vol. 26, no. 3, pp. 339–346, 2005.

[13] D. Anguelov, D. Koller, P. Srinivasan, S. Thrun, H.-C. Pang, and J. Davis, "The correlated correspondence algorithm for unsupervised registration of nonrigid surfaces," in *Advances in Neural Information Processing Systems*, Vancouver, Canada, 2004.

[14] M. van Eede, D. Macrini, A. Telea, C. Sminchisescu, and S.Dickinson, "Canonical skeletons for shape matching," in

Figure 6. Some incorrect classifications using the MNIST dataset produced by our approach, using 5-NN. As in [8], on the top of each handwritten digit, we show its identification, followed by its correct classification and the incorrect result produced by our algorithm.

Proc. of the Int. Conf. on Pattern Recognition. Washington, DC, USA: IEEE Computer Society, 2006, pp. 64–69.

[15] P. F. Felzenszwalb, "Representation and detection of deformable shapes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 2, pp. 208–220, 2005.

[16] H. Ling and D. W. Jacobs, "Shape classification using the inner-distance," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 2, pp. 286–299, 2007.