Convex Quadratic Programming for Object Localization

Hao Jiang, Mark S. Drew and Ze-Nian Li
School of Computing Science, Simon Fraser University, Burnaby BC, Canada V5A 1S6

Abstract

We set out an object localization scheme based on a convex programming matching method. The proposed approach is designed to match general objects, especially objects with very little texture, and in strong background clutter; traditional methods have great difficulty in such situations. We propose a convex quadratic programming (CQP) relaxation method to solve the problem more robustly. The CQP relaxation uses a small number of basis points to represent the target point space and therefore can be used in very large scale matching problems. We further propose a successive convexification scheme to improve the matching accuracy. Scale and rotation estimation is integrated as well so that the proposed scheme can be applied to general conditions. Experiments show very promising results for the proposed method in object localization applications.

1 Introduction

Finding objects in images, based on one or several template images, has been an important task in computer vision. Most early work on object location tries to locate objects by matching a cloud of points. In such a problem setting, the General Hough Transform and Geometric Hashing have been widely applied. Deterministic annealing schemes [1] have also been studied to match point clouds using a thin plate spline model. In recent years, matching based on invariant distinguishable features such as SIFT [2] and affine invariant features [3] has been attracting much attention. Invariant features have been successfully applied to matching texture-abundant objects. For objects that deform arbitrarily or have very little texture, invariant features are not very effective. In this paper, we study object localization based on both robust features and robust matching schemes.

Object localization based on features on a template is inherently a consistent labeling problem, which is NP-hard in general. A large volume of work has been devoted to solving the matching problem more efficiently. Apart from several special cases in which an exact solution is available by using dynamic programming [4] or maximum flow [5], which have polynomial time complexity, most general matching methods use approximation schemes. Greedy schemes such as ICM [6] have been studied for matching when features are relatively distinguishable. Graph Cut [7] and BP [8] are more robust matching schemes. Graph Cut has been mostly applied to stereo and BP has been applied to stereo and object matching applications.

In this paper we set out a straightforward approach by optimizing a consistent labeling problem. As in most consistent labeling formulations, our energy function has two terms: one matching cost term, and one smoothing term to enforce the consistent matching of nearby feature points. In this paper, we consider problems whose smoothing term is convex and can be represented by an L_2 norm. An L_1 relaxation scheme [9, 10] with city block distance has been studied by Jiang et al. in motion estimation and tracking. Here, we use a Euclidean Distance, in an L_2 relaxation for object localization applications. The successive convexification in [10] is in fact quite general and can also be applied to the L_2 relaxation. An L_2 formulation is more suited for applications when the displacement field is relatively smooth, which is true for most object localization applications. Using an interior point method, the L_2 formulation can be solved as efficiently as can the L_1 scheme.

Differently from most other approaches, we convert a non-convex consistent labeling problem into a sequence of easier convex quadratic programs. The convex QP relaxation is obtained by approximating the non-convex matching cost function for each template feature point with its lower convex hull. This scheme enables us to use a small number of basis target points to represent the whole target point space in the optimization process, and thus makes the scheme well suited for large label-set matching problems. To refine the matching, we propose shrinking the trust region for each feature point and at each stage solving a corresponding CQP. The proposed scheme is found to be able to almost always find the global optimum. Detecting scale and rotation is further integrated in the proposed scheme. Experiments show that this matching scheme is very robust and can solve problems for which invariant features become too sparse.

2 Object Localization based on Convex Quadratic Programming

2.1 Features for Matching

The features we use are the log-polar transformed image patches centered on the feature points in the target and template images. The log-polar transform simulates the human visual system's foveate property and puts more focus in the center view than the periphery views. The log-polar feature has large context and it is also not sensitive to small translations; this enables us to choose feature points more sparsely in the target image. Notwithstanding the fact that the log-polar transform feature increases robustness, matching is

The 18th International Conference on Pattern Recognition (ICPR'06)
0-7695-2521-0/06 $20.00 © 2006 IEEE

still very likely to fail without a robust matching scheme such as we present here.

2.2 Consistent Matching

Finding objects in an image can be formulated as a consistent labeling problem, in which we assign a corresponding point in the target image to a point on the template. The assignment or matching should make the cost of matching corresponding features low, and at the same time the assignment for neighboring template points should be consistent. We wish to find a mapping $\mathbf{f}$ from template points to the target points that optimizes the following problem:

$$\min_{\mathbf{f}} \left\{ \sum_{\mathbf{s}\in S} C(\mathbf{s}, \mathbf{f_s}) + \lambda \sum_{\{\mathbf{p},\mathbf{q}\}\in\mathcal{N}} ||(\mathbf{f_p} - \mathbf{p}) - (\mathbf{f_q} - \mathbf{q})||^2 \right\}$$

in which $C(\mathbf{s}, \mathbf{f_s})$ is the cost of matching point $\mathbf{s}$ on the template to the point $\mathbf{f_s}$ in a target image; S is the set of feature points on the template; $\mathcal{N}$ is the set of all neighboring point pairs on the template, which consists of all the point pairs connected by edges in the Delaunay graph of points in S; $||\cdot||$ is an L_2 norm. The objective function consists of two terms: the first is the matching cost term and the second is a smoothing term to enforce consistency of matching for nearby feature points. The weight of the smoothing term is controlled by a constant λ. This optimization problem is usually highly non-convex because of the non-convexity of the matching cost function $C(\mathbf{s}, \mathbf{t})$ with respect to $\mathbf{t}$. Notice that this formulation is not scale and rotation invariant; we will consider how to estimate the scale and rotation in § 2.5.

2.3 Convex Quadratic Programming Relaxation

The non-convexity of the matching problem makes it hard to directly solve the problem in its original form. Here, we propose relaxing the hard problem into an easier to solve convex programming problem. For each $\mathbf{s}$ on the template, we replace the matching cost $C(\mathbf{s}, \mathbf{f_s})$ with a linear combination of the *basis target point* costs: $C(\mathbf{s}, \mathbf{f_s}) \simeq \sum_{\mathbf{t}\in B_\mathbf{s}} \xi_{\mathbf{s},\mathbf{t}} \cdot C(\mathbf{s}, \mathbf{t})$, where $B_\mathbf{s}$ is the set of basis target points for $\mathbf{s}$ and $\xi_{\mathbf{s},\mathbf{t}}$ are non-negative coefficients. These points serve as a basis such that $\mathbf{f_s} = \sum_{\mathbf{t}\in B_\mathbf{s}} \xi_{\mathbf{s},\mathbf{t}} \cdot \mathbf{t}$, with the constraint $\sum_{\mathbf{t}\in B_\mathbf{s}} \xi_{\mathbf{s},\mathbf{t}} = 1$ for each $\mathbf{s}$. Then the relaxation problem becomes:

$$\min\{\sum_{\mathbf{s}\in S}\sum_{\mathbf{t}\in B_\mathbf{s}} \xi_{\mathbf{s},\mathbf{t}} C(\mathbf{s}, \mathbf{t}) + \lambda \sum_{\{\mathbf{p},\mathbf{q}\}\in\mathcal{N}} [(u_\mathbf{p} - x(\mathbf{p}) - u_\mathbf{q} + x(\mathbf{q}))^2 + (v_\mathbf{p} - y(\mathbf{p}) - v_\mathbf{q} + y(\mathbf{q}))^2]\}$$

with constraints:

$$\begin{aligned} \sum_{\mathbf{t}\in B_\mathbf{s}} \xi_{\mathbf{s},\mathbf{t}} &= 1, \forall \mathbf{s} \in S \\ u_\mathbf{s} &= \sum_{\mathbf{t}\in B_\mathbf{s}} \xi_{\mathbf{s},\mathbf{t}} \cdot x(\mathbf{t}), \\ v_\mathbf{s} &= \sum_{\mathbf{t}\in B_\mathbf{s}} \xi_{\mathbf{s},\mathbf{t}} \cdot y(\mathbf{t}), \forall \mathbf{s} \in S \\ \xi_{\mathbf{s},\mathbf{t}} &\geq 0, \forall \mathbf{s} \in S, \forall \mathbf{t} \in B_\mathbf{s} \end{aligned}$$

in which we denote functions $x(\mathbf{s})$ and $y(\mathbf{s})$ as extracting the x and y component of point $\mathbf{s}$. The matching result $\mathbf{f_s} \equiv (u_\mathbf{s}, v_\mathbf{s})$. Note that if we were to further constrain the variables $\xi_{\mathbf{s},\mathbf{t}}$ to be binary (0 or 1) and if $B_\mathbf{s}$ includes all the matching candidates for point $\mathbf{s}$, the optimization problem would be exactly equivalent to the original non-convex matching problem. But the integer quadratic program is hard to solve; we are most interested in the relaxed convex QP for which efficient solution schemes exist. The quadratic program has a close relation with the continuous extension of the non-convex matching problem: the continuous extension of a matching problem is defined by first interpolating the matching cost surface $C(\mathbf{s}, \mathbf{t})$ piecewise-linearly with respect to $\mathbf{t}$ and then relaxing feasible matching points into a continuous region (the convex hull of the basis target points $B_\mathbf{s}$).

The convex QP relaxation has several useful properties. When $B_\mathbf{s}$ contains all the matching candidates for $\mathbf{s}$, and the continuous extension cost function $C(\mathbf{s}, \mathbf{t})$ is convex with respect to $\mathbf{t}$, $\forall \mathbf{s} \in S$, CQP exactly solves the continuous extension of the discrete matching problem. In practice, the cost function $C(\mathbf{s}, \mathbf{t})$ is usually highly non-convex with respect to $\mathbf{t}$ for each site $\mathbf{s}$. In this case, the quadratic programming formulation solves the continuous extension of the reformulated discrete matching problem, with $C(\mathbf{s}, \mathbf{t})$ replaced by its lower convex hull for each site $\mathbf{s}$. In fact, we need only include the basis set $B_\mathbf{s}$ comprised of the vertex coordinates of the lower convex hull of $C(\mathbf{s}, \mathbf{t})$, $\forall \mathbf{s} \in S$, into the optimization process. The basis set has a much smaller number than that of candidate labels for one site, which thus greatly reduces the complexity for matching large label set problems. In a 2.6GHz PC, matching 50 feature points to 1000 candidate target feature points with CQP typically takes about 0.02 seconds.

2.4 Successive Convexification

A single relaxation is usually not sufficient to capture the non-convex details of the original optimization problem. Instead of just using one step of relaxation, we follow the successive convexification scheme [10] and successively convexify the matching cost surfaces in each step.

We define a trust region for each feature point on the template. Initially, the trust region for each feature point covers the entire target image. We then follow the proposed relaxation scheme to solve a convex QP and obtain an initial estimation. Based on the estimation, we can shrink the trust region of each site. In the new trust regions, the lower convex hull of the cost surfaces may change and we need to re-convexify the *original* surfaces and solve a new CQP. Such a process iterates until the trust regions become small. Successive convexification is illustrated in the gray area of the system diagram in Fig. 1.

In trust region shrinking, we use control points to anchor trust regions for the next iteration. We keep the control point in the new trust region for each site and we shrink the

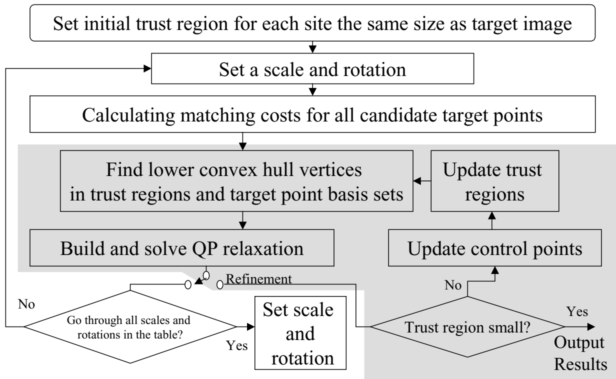

Figure 1. Object localization with CQP relaxation and successive convexification.

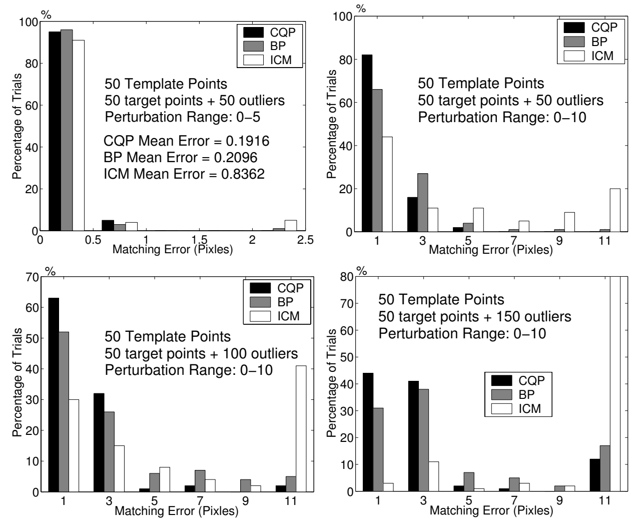

Figure 2. Matching random patterns.

boundary inwards. If the control point is on the boundary of the previous trust region, the other boundaries are moved inwards. We select control points using a consistent rounding process. In consistent rounding, we choose a site randomly and check all the possible discrete target points and select the one that minimizes the objective function in § 2.2, by fixing other sites' targets as the current stage CQP solution. This step is similar to a single iteration of an ICM algorithm by using CQP solution as initial value. We also require that new control points have energy not greater than the previous estimation. Such a trust region refinement scheme can greatly improve the matching result. The trust region usually shrinks fast in real applications. Typical iterations are 3 to 5.

2.5 Estimation of Scale and Rotation

Local features themselves do not provide enough information for scale of an object, and a consistent matching is necessary in scale and rotation estimation. We estimate the scale and rotation based on CQP with the largest trust region for each site. Since the template deforms, we can quantize the scale and rotation quite coarsely. The quantization levels for scale are 0.5, 0.75, 1, 1.25 and 1.5. The rotation is quantized with 45-degree intervals in 360 degrees. We then scale and rotate the template and obtain a matching with CQP relaxation. The matching score is the matching cost term $\sum_{\mathbf{s}\in S}\sum_{\mathbf{t}\in B_{\mathbf{s}}}\xi^*_{\mathbf{s},\mathbf{t}}C(\mathbf{s},\mathbf{t})$ in CQP's objective function, where $\xi^*_{\mathbf{s},\mathbf{t}}$ are optimal weights. Fig. 1 illustrates the matching process with scale and rotation estimation.

3 Experiment Results

In the first experiment we compare the convex QP method with BP and greedy scheme ICM using synthetic grayscale images. The template is a randomly generated image which has resolution of 128×128. We randomly place 50 white dots into the black background. The truncated distance transform of the binary image is then used as the template image. These dots are then randomly translated and perturbed and placed into another 256×256 image. Outlier points are added to the target image to simulate background clutter. This image then undergoes another transform similar to that of the template image and is used as the match target. In this experiment, the feature points are the 50 points placed in the black background; target candidate points are all target points including the outliers placed in the target image. The features used are log-polar transform image patches centered on the template and target points with masks of diameter 32. In each outlier and perturbation setting, we randomly generated 100 pairs of testing images and compared the results. Fig. 2 shows the matching error distribution for different methods, in several outlier and perturbation settings. In low distortion and low outlier cases, all three methods have similar performance. But when the outliers and distortion increase, greedy schemes degrade rapidly. The proposed scheme works the best in this experiment.

An object localization result using real images is shown in Fig. 3. The leaf used is an object with very little texture. Therefore the boundary edges of the leaf are the features that are used to locate the object. Because of the background changes from the template image to target image, these boundary features are also distorted considerably, which makes invariant feature based schemes fail to find the correct corresponding points. We use about 100 randomly selected edge points on the template and 1000 randomly selected edge points on the target image, as shown in Figs. 3 (a) and (b). Log-polar features are then used in matching. The matching scores for discretized scales and rotations are shown in Fig. 4 (c). The minimum cost rotation and scale estimations are correctly 45 degrees and 1.25 respectively. Successive convexification QP is then applied to locate the object in the target image. Fig. 3 (d) is the initial matching and Fig. 3 (e) shows the final matching result. Other object localization results are shown in Fig. 4, Fig. 5 and Fig. 6. In the experiments locating the toy and the hand in images, we use a smaller feature context and denser point candidates (about 8000 target points) in each target image, to increase the reliability of matching in complex backgrounds. BP becomes quite slow for such a large number of target points due to its $O(n^2)$ complexity with respect to the number of target points. ICM is not able to locate the targets correctly. For such large label-set matching problems, the proposed CQP scheme can efficiently locate the target objects in seconds, using an automatically generated template mesh.

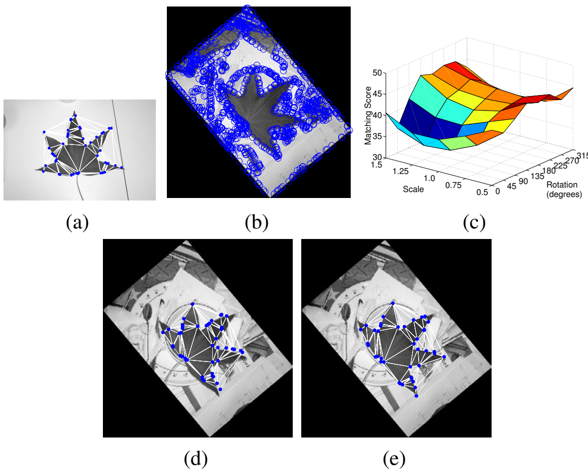

Figure 3. Leaf. (a): Template image and mesh; (b): Feature points on the target image; (c): Matching scores for different scales and rotations; (d): CQP matching in the largest trust region; (e): Final matching result.

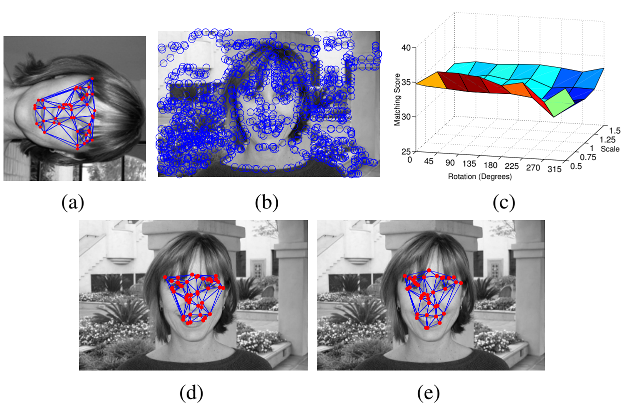

Figure 4. Face. (a): Template image and mesh; (b): Feature points on the target image; (c): Matching scores for different scales and rotations; (d): CQP matching in the largest trust region; (e): Final matching result.

4 Conclusion

We have set out an object localization method which can deal with textureless objects in strong background clutter. Finding objects in such situations is a challenging task for traditional invariant feature based schemes, which work well for texture-abundant objects. We propose a novel quadratic programming method to solve the class of hard, non-convex matching problems. The convex QP relaxation involves a very small number of basis target points in the search process and thus this method is well suited for very large label-set problems. We further propose a successive convexification scheme to refine the solution by systematically shrinking the trust region. The successive convexification QP has a high probability of converging to the global optimum. Experiments show very promising results for using this method in object localization problems. The matching scheme can also be applied to other problems such as motion estimation, tracking, and object recognition.

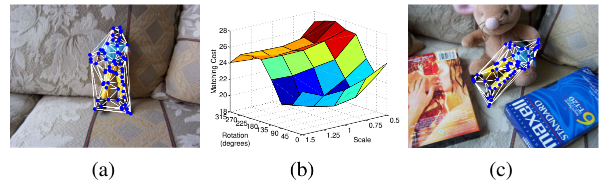

Figure 5. Toy. (a): Template image and mesh; (b): Matching scores for different scale and rotations. (c): CQP matching result.

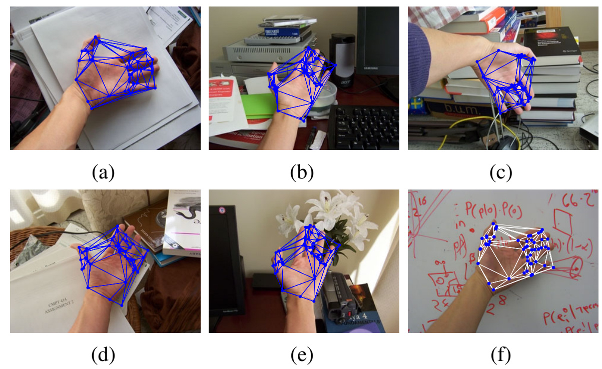

Figure 6. Hand. (a): Template image and mesh; (b)-(f): Hand localization results with convex QP.

References

[1] H. Chui and A. Rangarajan, "A new algorithm for non-rigid point matching", CVPR, vol. 2, pp.44-51, 2000.

[2] David G. Lowe, "Distinctive image features from scale-invariant keypoints", International Journal of Computer Vision, 60, 2 (2004), pp. 91-110.

[3] K. Mikolajczyk and C. Schmid, "Scale and affine invariant interest point detectors", IJCV, Vol. 60, No. 1, 2004.

[4] P.F. Felzenszwalb, "Representation and detection of deformable shapes", CVPR, vol.1, pp.102-108, 2003.

[5] H. Ishikawa, Global Optimization using Embedded Graphs, Ph.D. Dissertation. May 2000, NYU.

[6] J. Besag, "On the statistical analysis of dirty pictures", J.R. Statis. Soc. Lond. B, 1986, Vol.48, pp. 259-302.

[7] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts", PAMI, Vol.23, pp. 1222-1239, 2001.

[8] Y. Weiss and W.T. Freeman. "On the optimality of solutions of the max-product belief propagation algorithm in arbitrary graphs", IEEE Trans. on Information Theory, 47(2):723-735, 2001.

[9] H. Jiang, Z.N. Li, and M.S. Drew, "Optimizing motion estimation with linear programming and detail-preserving variational method", CVPR, pp.I:738-745, 2004.

[10] H. Jiang, M.S. Drew, and Z.N. Li, "Linear programming matching and appearance-adaptive object tracking ", EMMCVPR, LNCS 3757, pp.203-219, 2005.

0-7695-2521-0/06 $20.00 © 2006 IEEE