

Algorithms, Architecture, Validation of an Open Source Toolkit for segmenting CT Lung lesions

Karthik Krishnan¹, Luis Ibanez¹, Wesley D. Turner¹, and Ricardo S. Avila¹

Kitware Inc., 28, Corporate Drive, Clifton Park, NY, USA.

`karthik.krishnan@kitware.com`,

Abstract. We implemented a lesion segmentation toolkit with a focus on segmenting solid and part-solid lesions from lung CT. The algorithm constructs four three-dimensional features to detect the lung wall, vessels, soft and diffuse tissue and edges. These features form boundaries and propagation zones that guide the evolution of a subsequent shape detection level set. User input is used to determine an initial seed point for the level set and users may also set a region of interest around the lesion. The methods are validated against 18 nodules on an anthropomorphic thorax phantom simulating a realistic anatomy, acquired under various scanner parameters. We also validated repeatability using 6 coffee-break (zero volume change) clinical cases. The methods, clinical and testing datasets have been made available to encourage reproducibility and foster scientific exchange. We integrated the methods into a user-friendly visualization application. We will also describe how one may obtain the toolkit and generate results for the VOLCANO workshop.

1 Introduction

Computed Tomography (CT) is the preferred modality for the detection of cancerous lung lesions. Quantifying tumor size and changes in size over time is necessary to evaluate the effectiveness of a drug. Tumor size is currently quantified using the uni-dimensional RECIST [1] [2] or bi-dimensional WHO [2] [3], both of which use lesion diameter as surrogates for volume to determine an effective tumor measurement. Neither of these methods take the actual volume of the tumor into account. With the advent of MSCT, volumetric segmentation of lung nodules allows for more accurate quantification of the tumor burden and better determination of changes in tumor burden over time and in response to drug regimen.

The goal of this paper is to describe the algorithm and the architecture of an open source lesion sizing toolkit. We will validate it on phantoms as well as clinical data. We will also describe how one may obtain the toolkit and generate results for the VOLCANO [4] workshop. The rest of the paper is organized as follows. Section 2 summarizes the segmentation methods employed. Section 3 summarizes the validation of the methods on synthetic phantoms having known volumetric ground truth from the FDA. Section 4 outlines the architecture of the toolkit and describes the use of the toolkit from VolView, a commercially

available data visualization application. Integration with VolView allows users of the toolkit to load and visualize datasets as well as to perform volumetric analysis in an intuitive and powerful graphical user interface. Finally, in Section 5 we summarize our performance on the Workshop evaluation dataset and provide the user with those resources necessary to duplicate our results.

2 Lesion Segmentation

Lung nodules exhibit a wide range of morphologies and shapes. They may be solid, part-solid (solid nodules with a diffuse region) or non-solid. Solid nodules have a density that is close to soft tissue while non-solid nodules have a density between air and soft tissue. Nodules are generally attached to vessels and often to the lung wall or the mediastinum. Coding a single algorithm that effectively segments a nodule under these conditions is difficult. We choose to take a modular and extensible approach that combines a region growing algorithm - in this case a shape detection level set - with a set of feature generation filters. The region growing algorithm is allowed to evolve from a user provided seed under the aggregate guidance of the features. Our current implementation uses 4 features; to account for the lung wall, vessels, gradients and tissue intensities (soft and diffuse). Together, these features allow propagation of the segmentation front within the intensity thresholds specified by the intensity feature while deterring propagation into the lung wall, into vessels, or across sharp edges. The addition of new features to specifically account for areas such as eg. the hilar region is easily supported.

2.1 Feature generation

Prior to any feature generation, the CT dataset is preprocessed by resampling to isotropic resolution matching the smallest voxel spacing. This is followed by generation of various features:

Lung wall feature The lung wall is segmented by thresholding followed by curvature constrained front propagation. The dataset is binarized so as to retain voxels above -400HU. We then run hole filling on the resulting binary image. The hole filling operation uses a quorum (voting) algorithm to decide whether a new pixel will be *filled* at every iteration. Changing the voting threshold is essentially equivalent to changing the maximum curvature allowed in the final binary segmentation. We use a $7 \times 7 \times 7$ manhattan kernel with a voting threshold of 1 (lowest possible curvature). The filter is run iteratively until no voxel changes status. Fig.1 shows the evolution of the front. Admittedly, the final contour is slightly concave rather than convex, since the quorum voting does not distinguish between positive and negative curvature, but the wall detection recovers most of the nodule without allowing bleeding into the lung wall.

The resulting image is transformed using a sigmoid with inversion, resulting in a grayscale representation of the inner lung, with a gradual rolloff towards the lung wall.

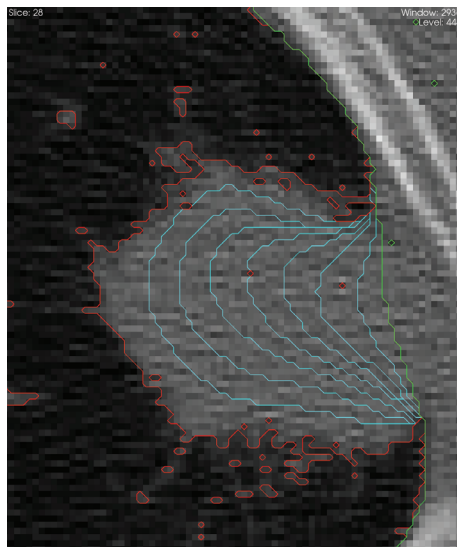


Fig. 1. Segmentation of the lung wall by voting based front propagation on a lesion attached to the lung wall. The front starts with the red contour (obtained by thresholding the dataset at -400HU). In cyan are contours of the front at iterations of 10, 20, 30, 40, 50, 60, 80. The algorithm converges in 151 iterations at the green front.

Vesselness feature Vesselness features are computed using Sato [5] tubularness. ($\alpha = 0.1$, $\beta = 2.0$). The resulting image is transformed using a sigmoid filter with inversion resulting in a grayscale image with a rolloff towards tubular structures.

Gradient feature The gradient feature attempts to localize the edges of the lesion. This is modeled using a Canny edge detector [6], which achieves optimal edge localization (with voxel accuracy). The result is a voxelized edge map. The resulting image is transformed using a sigmoid with inversion resulting in a grayscale image with a rolloff towards the edges.

Intensity feature The intensity feature is based on a simple thresholding of the dataset, followed by transformation with a sigmoid filter to generate gradual rolloffs. The user specifies the type of the lesion (solid / part-solid), which determines the threshold level: -200HU for solid lesions and -500HU for part-solid lesions.

Aggregation of features Conceptually, each feature image represents the likelihood of a particular voxel being part of the lesion from a single perspective. Fig 2 and 4 show the features on a solid and a part-solid lesion. An aggregation

Algorithm 1 Voting based front propagation

```

1: Let Kernel Radius  $r = 3$ 
2: Let BirthThreshold,  $T = r^3 + 1$ . For a  $7 \times 7 \times 7$  city-block kernel,  $T = 171$ 
3: Let  $A_i$  be indices of voxels on the front at iteration  $i$ 
4: Threshold CT image at -400HU. Voxels above are set to 1, below to 0.
5: Iterate over the image and add all indices within  $r$  voxels of the boundary to  $A_0$ 
6: repeat
7:    $PixelsChangingStatus \leftarrow 0$ 
8:   for each voxel in  $A_i$  do
9:     Check for quorum,  $Q$ . (Quorum is achieved if number of ON voxels within a
       neighborhood ( $7 \times 7 \times 7$ ) centered at the current voxel  $> T$ .
10:    if  $Q$  is true then
11:      Schedule this pixel for inclusion in foreground.
12:       $++PixelsChangingStatus$ 
13:      Add background voxels in neighborhood to front for next iteration,  $A_{i+1}$ .
14:    else
15:      Schedule this pixel for inclusion in background.
16:      Add this location to the front for the next iteration,  $A_{i+1}$ 
17:    end if
18:  end for
19:  Update the status of scheduled voxels.
20:   $++i$ 
21: until  $PixelsChangingStatus = 0$ 

```

of the features is obtained by normalizing each individual feature to lie within the range 0 to 1 and then choosing the minimum feature value at each voxel. The resulting image is passed onto the segmentation module.

2.2 Segmentation Module

The segmentation module runs a segmentation algorithm guided by a feature probability image. The toolkit is equipped with a variety of segmentation modules, wrapped in a way that allows them to be plugged in as components to the lesion segmenter. In our user interface, the user identifies the lesion with a single seed point. We use fast marching [7], followed by refinement with a shape detection level set [8]. The fast marching front starts at the seed and proceeds outwards, solving the eikonal equation, its propagation guided solely by the aggregate feature image and an optional, user-defined, region of interest (ROI). The output is a time crossing map, with each voxel representing the time taken for the front to reach that voxel. We use a stopping time of 5s implying that the front stops propagating when all voxels on the front have an arrival time greater than 5. The fast marching front is conservative enough to ensure that the segmentation results lie within the lesion. This resulting time crossing image is renormalized to the range $[-4, 4]$, so as to be symmetric about 0 and passed onto the shape detection module.

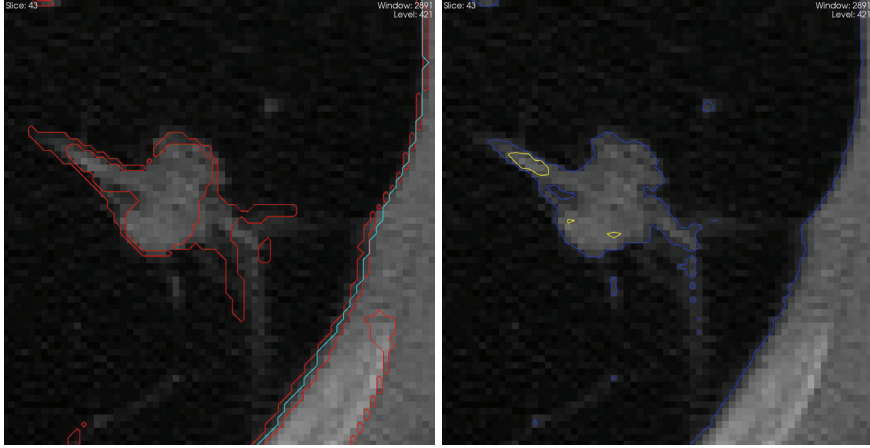


Fig. 2. Features (a) Lung wall features in cyan, Canny Edges features in red; a contour around the single voxel thick edges is shown (b) Intensity features in blue, vesseness features in yellow.

The shape detection level set is guided by the propagation and curvature components.

$$\phi_t + F(1 - wK) |\nabla\phi| = 0$$

The propagation F is the aggregate feature image. w is the curvature weight for the curvature, K . We use a propagation to curvature weight ratio of 500:1, to account for the wide variety of shapes that lesions can assume. The final lesion boundary is then taken as the -0.5 isosurface of the shape detection level set. Note that this isosurface value is driven by the characteristics of our aggregate feature map. In the nominal case of the lesion having a boundary within the parenchyma, we find that the canny edge is the dominant feature for lesion segmentation. Our canny edge feature only localizes the true edge boundary to a one voxel precision (Fig.4(b)). By taking the level set at -0.5 we force the segmentation to the interior of the canny detected boundary voxel and reduce our error bound from 1 to 0.5 voxels.

A surface of the resulting level set is generated using Marching Cubes [9] at an isovalue of -0.5. The volume is computed using a discretized version of the divergence theorem [10], to compute the volume in a closed triangle mesh.

3 Validation of the segmentation algorithm

3.1 Phantoms and Ground truths

We validated the results on an anthropomorphic thorax phantom [11][12] (courtesy US FDA), containing nodules of various densities, sizes, shapes and at various locations in the chest. This phantom provides ground truth for nodules

in a realistic anatomy (fig. 3) that includes attachments to vasculature, airways and the lung wall. Exposure varies from 20mAs to 200mAs. At the time of this writing, ground truth volumes are available only for the spherical nodules. We are awaiting ground truths via microCT on the spiculated, lobulated and elliptical nodules. The database consists of 6 CT scans, each containing a 5, 8 and 10mm diameter spherical nodule with a density of 100HU. Three of the scans have a slice thickness of 0.8mm and a z spacing of 0.4mm, the other three have a slice thickness of 3mm and a z spacing of 1.5mm.

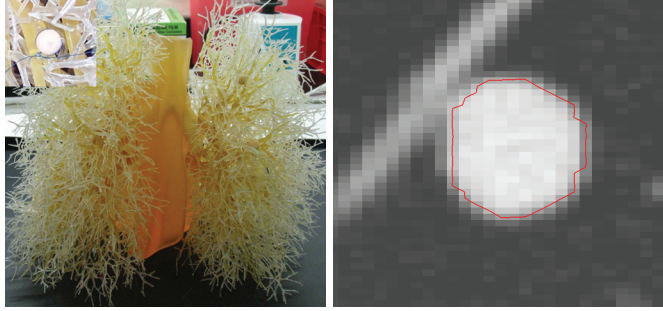


Fig. 3. *Left:* A photograph of the anthropomorphic phantom and an inlaid nodule (courtesy US FDA). *Right:* Its segmentation.

3.2 Statistics

Defining the error, $\xi = 100 \times \left| \frac{\text{ComputedVolume} - \text{TrueVolume}}{\text{TrueVolume}} \right|$, we find that for the 18 lesions in the FDA phantom, our average ξ was 30.24% (S.D 31.39%). The errors largely stemmed from a poor estimate of the features due to the resolution across slices. This was confirmed by noting that on the 9 lesions, with 0.8mm slice thickness, our average volumetric error, ξ was 11.75% (S.D 11.38%). Furthermore, the half voxel error (as mentioned above) causes an amplification of the errors for the small 5mm lesions. This is confirmed by noting that on the 6 lesions with a diameter $\geq 8\text{mm}$ and a slice thickness of 0.8mm, our average volumetric error ξ is 4.28% (S.D. 2.4%).

It is instructive to relate the error in volumetric estimate to the error in the uni-dimensional RECIST measurement. A 30% overestimate in a lesion with a diameter of 8mm is equivalent to a change in diameter of 0.73mm. Hence a ξ of 30% for a 8mm lesion would roughly which roughly translates to an error of a voxel. This is smaller than the intra as well as interobserver variability to perform a unidimensional tumor measurement [2].

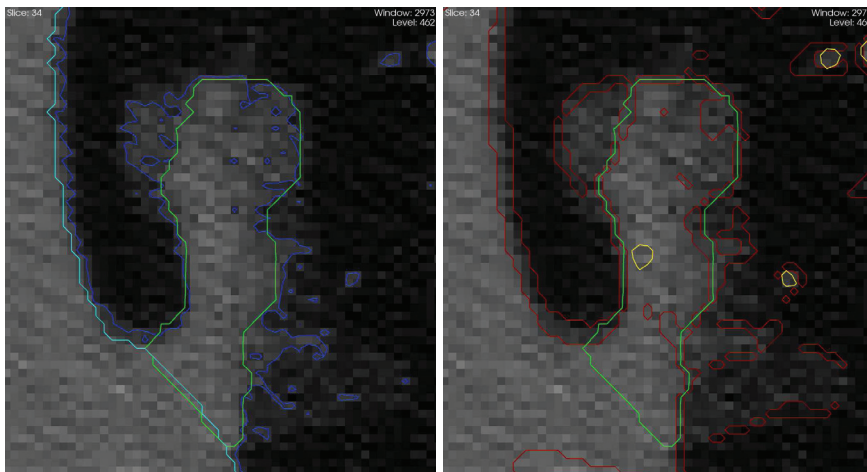


Fig. 4. Features (a) Lung wall feature in cyan, Intensity features in blue. (b) Vesselness features in yellow, Voxelized canny edge features in red, final segmentation in green. Note that the the final segmentation runs through the center of the voxelized canny edges as a result of cutting the level set at -0.5.

3.3 Zero change CT scans

Validation on clinical data was done on 6 coffee-break CT scans. (Table 1) These are CT scans of a patient obtained hours apart. (hence the name coffee-break). Ideally, segmentation would yield no change between the timepoints. However breathing artifacts, subtle differences in patient positioning, and noise result in slightly different segmentations. In addition, the two timepoints often had different acquisition parameters or resolution. The average ξ for the coffee break data was 10.9%.

Table 1. Zero Change Volumes V_1 , V_2 for the 6 coffee break cases at two time points.

Case	$V_1 mm^3$	$V_2 mm^3$	Resolution (t1)	Resolution (t2)	$\xi = 200 \times \frac{V_1 - V_2}{V_1 + V_2}$
ST0108	1992.76	1990.41	$.703 \times .703 \times 1.25$	$.703 \times .703 \times 1.25$	0.11%
ST0109	2390.58	2698.9	$.703 \times .703 \times 1.25$	$.703 \times .703 \times 1.25$	12.11%
ST0111	3496.34	2484.37	$.703 \times .703 \times 1.25$	$.703 \times .703 \times 1.25$	33.84%
ST0112	290.07	297.62	$.703 \times .703 \times 1.25$	$.703 \times .703 \times 1.25$	2.56%
ST0113	670.27	654.42	$.703 \times .703 \times 1.25$	$.78 \times .78 \times 1.25$	2.39%
ST0114	3317.62	2864.81	$.56 \times .56 \times 1.25$	$.56 \times .56 \times 5$	14.64%

4 Toolkit and User Interface

The toolkit [13] is provided under a BSD license. It is cross platform and based on the Insight [14] and Visualization Toolkits [15]. Fig. 5 shows the general architecture of the toolkit. Several feature generators and segmentation modules are provided, with an applicability towards a broader class of segmentation applications. A dashboard tests the algorithms against the aforementioned database providing a testbed to evaluate segmentation algorithms.

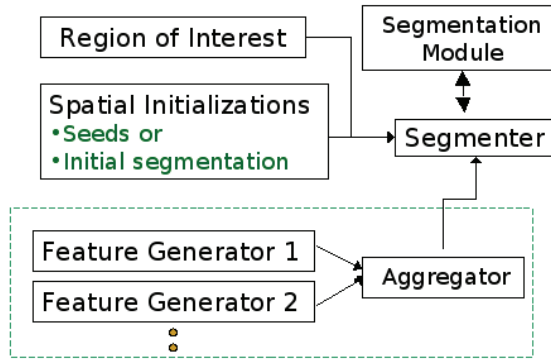


Fig. 5. Toolkit architecture.

4.1 Front end user interface

VolView [16] is visualization tool from Kitware. It provides a front end to the toolkit and features a guided wizard to perform one click segmentations of solid/part-solid lesions. The user identifies the region containing the lesion with a box, places a seed within the lesion and selects the lesion type (solid/part-solid). The application then segments the lesion and reports volumetrics. (Fig. 6.) A typical segmentation takes 5 to 25 seconds on a Intel 2.66GHz Core2Duo, 2GB RAM.

5 Results on the VOLCANO Grand Challenge data

For the VOLCANO Grand Challenge, all parameters remain as described above. No additional training was used to tailor the algorithm beyond the phantom data described in Section 3. User interaction consisted of the seed points and an ROI synthesized as a box 50mm on each side centered at the seed pixel.

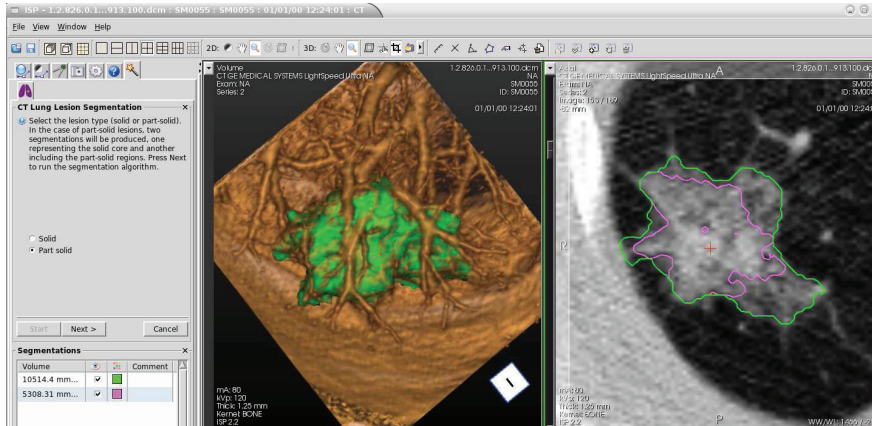


Fig. 6. A screenshot of the lesion sizing user-interface in VolView, on a part-solid lesion. The solid core (pink) and part-solid fringes (green) are segmented.

5.1 Generating the results

We provide the ability to automatically build and run the application against the Grand Challenge data. The user is expected to download the datasets from the VOLCANO [4] website. Instructions on obtaining the toolkit and building it may be found at [13]. Configure with the TEST_VOLCANO_DATA_COLLECTION option turned on. Build and run the tests. This will produce the volumetric estimates and segmentations for each of the datasets in the evaluation database.

5.2 Results

The results on the evaluation dataset are summarized in Table 2. Change is evaluated as $(V2 - V1)/V1$.

6 Conclusions

We have presented an algorithm to segment lung lesions from CT. We validated the results on clinical data and phantoms that provide a realistic anatomy. An important component of this work is reproducibility through open source, sharing of data and dashboards. Future directions may incorporate the use of scanner parameters. We will also validate against the spiculated and lobulated lesions in the anthropomorphic phantom, once ground truths are available.

7 Acknowledgements

We thank the Optical Society of America and the Air Force Research Laboratory for funding this work. We also thank David Yankelevitz (Weill Cornell Medical

Center) and Anthony Reeves (Cornell University) for the clinical and zero change cases. We also thank Nick Petrick and Lisa Kinnard, Marios Gavrielides and Kyle Myers from the FDA. The lung phantom data was provided by the U.S. Food and Drug Administration CDRH/OSEL/DIAM. The FDA collection was partially supported with funding through the intramural programs of NCI and NIBIB from the FDA for the phantoms.

References

1. Therasse, P., et al: New guidelines to evaluate the response to treatment in solid tumors. *J. National Cancer Institute* **92**(3) (Feb 2000) 205–16
2. Erasmus, J.J., Gladish, G.W., Broemeling, L., Sabloff, B.S., Truong, M.T., Herbst, R.S., Munden, R.F.: Interobserver and intraobserver variability in measurement of nonsmall-cell carcinoma lung lesions: Implications for assessment of tumor response. *J. Clinical Oncology* **21** (Jul 2003) 2574–82
3. : WHO handbook for reporting results of cancer treatment. (1979)
4. : <http://www.via.cornell.edu/challenge/>.
5. Sato, Y., et al: 3D Multi-Scale line filter for segmentation and visualization of curvilinear structures in medical images. *CVRMed MRCAS* (1997) 213–22
6. Canny, J., F.: A computational approach to edge detection. *IEEE Trans. Pattern Analysis and Machine Intelligence* **8**(6) (1986) 679–698
7. Sethian, J.A.: *Level Set Methods and Fast Marching Methods*. Cambridge Press (1999)
8. Malladi, R., Sethian, J., Vemuri, B.C.: Shape modeling with front propagation: A level set approach. *IEEE Trans. Pattern Anal. Machine Intel.* **17**(2) (1995) 158–75
9. Lorensen, W.E., Cline, H.E.: Marching cubes: A high resolution 3d surface construction algorithm. In: *SIGGRAPH*. Volume 21. (July 1987) 163–169
10. Alyassin, A.M., Lancaster, J.L., Downs, J.H.: Evaluation of new algorithms for interactive measurement of surface area and volume. *Medical Physics* **21**(6) (1994)
11. Gavrielides, M.A., Zeng, R., Kinnard, L.M., Myers, K.J., Petrick, N.A.: A matched filter approach for the analysis of lung nodules in a volumetric CT phantom study. *Proc. SPIE Med. Imaging* (Feb 2009)
12. Gavrielides, M., Kinnard, L., Park, S., Kyprianou, I., Gallas, B., Badano, A., Petrick, N., Myers, K.: Quantitative in silico imaging and biomarker assessment using physical and computational phantoms: a review of new tools and methods available from the NIBIB/CDRH joint Laboratory for the Assessment of Medical Imaging Systems. *Radiology* (2008)
13. : Lesion sizing toolkit, <http://public.kitware.com/LesionSizingKit>.
14. Ibanez, L., Schroeder, W., Ng, L., Cates, J.: *ITK Software Guide*. Kitware Inc.
15. Schroeder, W., Martin, K., Lorensen, W.: *Visualization Toolkit*. Kitware Inc.
16. : Kitware Inc. Volview <http://www.kitware.com/VolView>.

Table 2. Results on the VOLCANO evaluation dataset

case	Change	V1 (mm^3)	V2 (mm^3)
SC0001	-0.93	4352.2	286.4
SC0002	-0.14	1942.94	1673.08
SC0003	-0.41	3805.55	2248.26
SC0004	0.55	3145.82	4873.41
SC0005	-0.12	2457.89	2161.53
SC0006	0.54	141.01	217.1
SC0007	0.01	393.48	398.88
SC0008	0.02	116.91	118.74
SC0009	-0.28	8029.24	5765.85
SC0010	-0.13	1187.27	1035.82
SC0011	0.06	1856.14	1964.2
SC0012	0.59	3016.81	4804.87
SC0013	0.04	16440.7	17177.9
SC0014	-0.11	8583.22	7635.9
SC0015	-0.17	400.99	331.64
SC0016	0.13	330.54	374.03
SC0017	1.05	1096.42	2252.93
SC0018	0.04	2289.68	2371.25
SC0019	0.54	2505.32	3861.29
SC0020	-0.02	1151.27	1125.71
SC0021	-0.01	1564.57	1551.21
SC0022	0.17	593.7	694.95
SC0023	-0.04	2172.98	2077.6
SC0024	0.14	820.46	938.44
SC0025	0.13	970.89	1094.68
SC0026	0.05	1493.97	1562.17
SC0027	0.13	12571.3	14195.3
SC0028	-0.06	1322.43	1244.15
SC0029	0.04	198.47	206.11
SC0030	-0.09	7628.96	6962.92
SC0031	-0.13	21902.6	19033.2
SC0032	-0.04	200.84	192.44
SC0033	0.1	481.6	528.31
SC0034	1.08	400.36	831.86
SC0035	-0.01	3021.52	2985.93
SC0036	0.08	892.11	966.66
SC0037	0.26	979.42	1236.19
SC0038	-0.02	1530.35	1506.32
SC0039	-0.33	1466.36	978.5
SC0040	-0.15	783.49	666.37
SC0041	-0.09	934.45	851.96
SC0042	0.12	584.12	652.95
SC0043	0.14	2338.12	2659.17
SC0044	1.46	438.45	1076.65
SC0045	0.03	678.82	697.13
SC0046	0.03	795.27	818.08
SC0047	0.13	1166.88	1317.97
SC0048	-0.25	511.91	384.87
SC0049	-0.35	4638.86	3004.46
SC0050	0.22	634.62	773.72