Pattern Constancy Demonstration of Retinex/Visual Servo (RVS) Image Processing

Smart Visual Awareness System (SVAS)
Subtask of External Hazard Detection
Integrated Intelligent Flight Deck Technology
Aviation Safety Program

Co-Investigators: Zia-ur Rahman[+] (Old Dominion University), Daniel J. Jobson [*] (NASA-LaRC), and Glenn A. Woodell (NASA-LaRC)

+Dr. Rahman's contribution was funded by
NASA Aviation Safety NRA#NNL07AA02A

*SVA Subtask Team Leader

The key ideas of this subtask are:

1) Previously developed real-time generic image enhancement methods, <u>retinex/visual servo (RVS) image processing</u>, were found to be approaching a canonical visual representation for arbitrary scenes and arbitrary imaging and sensor conditions - changing lighting and atmospheric conditions such as turbid imaging, and sensor exposure variations.

2) Retinex/visual servo methods needed to be optimized for maximum stability of this canonical representation and the canonical hypothesis itself needed to be demonstrated experimentally. Together, these research efforts provide the necessary foundation for the use of the RVS as a generic, enabling, computational front end for subsequent higher level visual processing which performs visual tasks such as pattern recognition. This combined processing is then directed to achieve automatic external hazard detection for future avionic systems that will increasingly rely on smart sensor/processor technology to augment the pilot’s situational awareness. The hazards amenable to visual information processing solutions are: runway and runway hazard detection, terrain hazard detection, and mid-collision hazard detection.

3) To demonstrate the canonical representation performance of the optimized RVS, we have studied a variety of time series image classes and carried them through RVS processing and then on to detected edge patterns to confirm that canonical visual representation leads to pattern constancy in the pattern domain of higher level processing. These experimental results yield a reasonably comprehensive understanding of the performance envelope which can be

accomplished, and scientific insights into the intrinsic limits on this performance.

4) Beyond the results here, we will be exploring both generic and specialized higher level processing methods which have the potential to achieve automated aviation hazard detection of the kinds listed above.

The RVS processing is described and its performance demonstrated as an automatic, powerful, and generic image enhancement method in previous publications (Ref. 2, 3, 4, 6, 11, 13-16, 19, 23-25 among many others in our collection of publications at http://dragon.larc.nasa.gov/retinex/background/retpubs.html) which include real-time video hardware implementations in digital signal processors (DSP). These later reports support the idea of this technology being readily embedded in future smart aviation imaging systems as a computational platform for higher level autonomous, external hazard detection visual processing. For the present purposes, the RVS processing, at its core, should be understood as a basic transformation of images from being a map of relative light intensity values to a map of the logarithm of spatial and spectral scene relationships- i.e. a compressed map of context information.

I. Approach

In order to build a comprehensive demonstration of canonical representation and pattern constancy, we examine time series imagery of the same scenes under widely varying imaging and sensor exposure conditions. Our intent is to demonstrate that any specific scene can be encountered again and again with widely changing extraneous variations and still produce a processed result which maintains a high degree of visual representation consistency and pattern constancy. Time series data sets that were readily at hand are:

1) <u>High quality webcam data sets</u>

   These data are from reasonably high quality web cams and are high quality lossy JPEG image formats. Noise will be seen to be the performance limiting factor except for the San Marino webcam where lossier JPEG coding introduces block artifacting as the pattern constancy limit rather than sensor noise. These data extend the Mars data set to terrestrial landscapes and are rich in temporal variations due to weather, lighting, and even seasonal variations. So where the Mars data is rich in locational and scene diversity, the web cams are rich in temporal depth but sparse in locational diversity. The locations included are Paris, France metropolitan panorama, San Marino, Italy (a mix of natural and city terrain on mountain ridge), and Vancouver BC, Canada downtown skyline with highly variable weather. Other sites also included are two Italian mountain village webcams where lighting variations are accentuated by mountain shadows.

2) <u>Mars Odyssey orbital imagery</u> of the surface of the planet

The key attributes of this data are:

1) Rich diversity of terrain features (and patterns)

2) Time series of two to six images per scene are available and represent the scene over highly arbitrary lighting and sensor exposure variations

3) The data are high quality lossless TIF, GIF, or PNG image formats. This data set is limited in the number of time series/location available since the primary mission of Mars Odyssey was planetary survey rather than the intensive study of selected locations on the surface. This data is also limited to essentially no weather or turbidity variations. So the data set is rich in locations but poor in time series for each location, and lacking in terrestrial weather variations. While image quality is limited by noise rather than lossy image coding artifacts, this noise is not always stationary and frequently can be seen to be afflicted with periodic variations.

3) Fog sequence at NASA LaRC Aircraft Landing Dynamics Facility (ALDF)

In order to augment data readily at hand, we also took a sequence of morning fog clearing images at Langley's ALDF. This gave us a well-characterized sequence of turbidity variation from image data we acquired and had full control over the image acquisition process. These image sequences were acquired with a Canon EOS-1 Mark II N digital camera.

Taken together, the above data sets comprise a reasonably comprehensive array of highly arbitrary scene diversity, imaging lighting and turbidity variations, and sensor exposure variations. The latter is especially emphatic for the Mars Odyssey where we often see images in a series taken at the same sun angle (judged by crater shadows) but with wildly different sensor exposures and resulting feature contrasts ranging from invisible to high contrast.

A basic summary description of all data sets is:

A) High quality web cams (high quality lossy JPEG image formats)

Italy-

1. Unknown mountain village- 1 location, 8 time series spanning one day/night cycle

2. Pellegrino- 1 location, 15 time series spanning one day/night cycle- bad weather

3. San Marino- 1 location, 40 time series over ~ 1 month- very little bad

weather- mostly sun angle variations in lighting

Vancouver BC, Canada- 1 location, 49 time series over ~ 1 month- lots of bad weather and highly variable weather

Paris, France- 1 location, ~50 time series over ~ 1 month- lots of weather variations- rain/fog/haze

In addition, we processed image data sets from some low quality webcams. These were severely limited in image quality and enhancement potential by the presence of emphatic block artifacting due to very lossy image coding. However we do show some data for Salzberg, Austria (castle/airport) with very poor visibility and highly variable lighting/fog conditions, because it does still show the processing demonstrably pushing toward canonical visual representation despite the very serious limits imposed by such lossy image coding.

B) Mars Odyssey- >1500 locations on Mars surface comprised of very rich terrain diversity- 2-6 time series for each location usually acquired years apart in time. Loseless TIF, GIF, or PNG image formats. The plethora of time series data available for the Mars surface allowed us to tease apart the extraneous variables so that we could examine the individual effects of sun angle, sensor exposure, and signal-to-noise variabilities. This separation of variables is shown in separate sub-galleries in the Mars Odyssey results gallery.

C) Fog sequence- Thirteen images acquired in a time sequence as morning fog was clearing.

II. Results (see galleries)

We have selected for the gallery, samples of this data set which represent typical performance as well as examples of cases at the extremes of performance, namely very low light levels, poor visibility turbidity, and very poor sensor exposures. The typical performance is usually highly redundant so we are not showing as many of these excellent performance cases as we could, but concentrating on the cases approaching performance limits in order to illustrate these limiting cases and develop insights into the mechanisms of these limits.

Discussion of results-

Both enhancement and canonical representation performance as well as pattern constancy performance are quite good over a wide range of extraneous imaging variabilities but at the extreme are limited by sensor noise (or lossy image coding artifacts if present). Lesser limits on constancy and other performances are variable shadow edges, loss of small detail in small shadows, and any losses due to signal saturation and clipping (such information is irretrievably lost). With respect to ultimate pattern recognition on the Mars pattern constancy data sets, we do see that certain scenes encountered on Mars are somewhat featureless or granular. This sort of textural pattern may lack sufficient structural uniqueness to support an unambiguous pattern recognition computation. Also in some rarer cases, more rolling terrains are less pattern-constant with sun angle changes, however the pattern constancy holds well for scene elements with high relief such as craters, canyons, and mountain ridges. On the whole, pathological cases of pattern inconstancy are very rarely occurring.

III. Summary of Visual Judgments of Pattern Constancy for the Very Diverse Mars Odyssey Data Sets:

The overall results of visual judgment for 369 of the >1500 cases are:

70% likely to strongly correlate
14% marginal correlation
16% poor to no correlation

These data are therefore for a large sample of the total data set.

Again, the reasons for the marginal to no correlation are poor s/n ratio, featureless or granular textural terrains, or scene change such as reorientation of dune pattern after a storm, or melting and formation of ice patterns obscuring terrain features in polar regions. That being said, the vast majority of the poorly matched cases are due to sensor noise limits- a fundamental and unavoidable limit for any imaging system.

Without image enhancement, these statistics would be heavily skewed to marginal or no correlation classes. So pattern constancy is demonstrated to a significant degree down to the fundamental limit of sensor noise.

IV. Visualizing Pattern Constancy (see gallery)

In order to give a more concrete sense of pattern constancy, we have included a gallery where time series pairs of detected edge patterns are overlaid with t_1 in blue and t_2 in red. The common pattern elements then appear magenta and can readily be distinguished from the non-matching edge elements which remain either blue or red and clearly designate from which of the non-matching edge element image pairs each comes from.

V. A Cursory Correlation Analysis of Time Series Data (see gallery)

While it is logistically impossible to precisely spatially register such a large volume of time series data as we have shown here in the galleries, we did want to carry the visual assessment of pattern constancy on to its computational pattern recognition impact. Therefore we ran a new terrestrial time series where camera position was absolute so that all data was intrinsically registered and correlation coefficients could readily be computed. In addition, we were also able to register some of the Paris panorama webcam images and the fog sequence shown earlier. The computation of a correlation coefficient gives us quantitative insight into the the performance of the enhancement processing and its ability to boost the likelihood of a valid pattern recognition event. The basic process which these data describe is that a good edge representation of a scene is stored in memory and the attempt is made to correlate this edge image with others coming into the processing, from real-time sensor data for example. We hope to also assess the effect of scene content and context on pattern constancy performance by using time series data from rather different scenes.

From experiments with increasing injections of noise in test images, we found that the noise limit for correlation is r=0.00-0.04 and for non-matching patterns range from r=0.00 to r=0.05. This supports the idea that a significant correlation value can be as low as ~0.1 or even less, perhaps down to 0.06. With this in mind, we can evaluate the three sets of time series correlation data.

A Residential Street Scene

In addition, we did a residential street scene correlation analysis of time series images ranging from near saturation through normal exposures all the way down to near darkness. This experiment and the results are given below.

For a full sequence of time series of images from a residential street scene, we computed the behavior of the correlation coefficient over a wide array of varying conditions- near saturation through normal exposures with variable lighting (sun angle mostly, down to advancing nightfall with increasing sensor noise to its extreme limits. The correlation coefficient is computed for both un-enhanced edge images and the enhanced edge images to quantify the benefit of enhancement, per se, and to quantify the degree of pattern constancy that is maintained to "recognize" this specific scene with a computer under these highly arbitrary imaging conditions.

The chronological order of the images has been rearranged to show the trend from near saturation down to deepening night. The best enhanced visual representation of the scene is chosen to define the edge image to correlate against all the other edge images- both un-enhanced and enhanced. **R is the correlation coefficient computed as the covariance divided by the product of the two separate standard deviations.**

Notes	File Name	R (un-enhanced)	R (enhanced)
almost complete saturation	house2-1ee	0.00000	0.04
decreasing saturation	house2-2ee	0.037	0.30
	house2-3ee	0.22	0.50
....	house2-4ee	0.43	0.62
	house2-5ee	0.49	0.72
near normal exposure	house2-6ee	0.48	0.79
best visual representation	house2-7ee	0.50	1.0

	house2-8ee	0.41		0.69	

**Major scene change (dumpster moved to block a major foreground pattern in previous scene)
Shifting to new correlation baseline image

twilight begins	house2-9ee	(0.16)	0.48	(0.17)	1.0
darkness deepens	house2-10ee		0.31		0.67
increasing noise	house2-19ee		0.089		0.43
	house2-18ee		0.037		0.35
......	house2-17ee		0.004		0.23
	house2-16ee		-0.003		0.14
	house2-15ee		-0.003		0.07
	house2-14ee		-0.003	0.03	
nearly all noise	house2-11ee		-0.001		0.02

NOTE: From previous experiments the noise limit for correlation is 0.0-0.04 and the mismatch detection limit is for R of 0.00-0.05. So for R>0.05 we seem to have a significant recognition or detection event. If so, then the RVS processing extended the “recognition” events from house2-2ee all through house2-15ee- that is, from very bright, near saturation to very deep twilight, almost black

B. Paris Panorama Webcam

While the zoom and view position of this camera did change somewhat over time, which made spatial registration unmanageable, we were able to register enough e xamples to cover a very wide range of lighting conditions including down to the noise limit at night. These data, like the previous set, are not arranged in chronological order, but rather to show the trend for progressive declines in enhanced image quality.

Notes	File Name	R (un-enhanced)	R (enhanced)
	paris02ee	0.57	1.0
	paris17ee	0.19	0.33
	paris40ee	0.24	0.32
	paris15ee	0.26	0.30
	paris18ee	0.15	0.27
	paris33ee	0.16	0.25
	paris05ee	0.14	0.23
extreme lighting*	paris38ee	0.08	0.23
	paris34ee	0.16	0.22
very high noise	paris01ee	0.03	0.03

*near-sunset glare condition

These data exhibit several interesting trends. Foremost is the lower enhanced R values compared to the previous residential street scene. A careful examination of the data showed that while there is some residual spatial mis-registration of these time series, most of the lower R values are due to scene context and content. Unlike the residential street scene, this panorama is composed of much more fine structure and as such is more subject to pattern inconstancy due to local shadow changes as a result. Such local shadow changes are not removed by the RVS processing. The lower correlation values reflect this and quantify the “core” pattern constancy achieved by the RVS processing for the class of scene. The RVS processing is clearly maintaining a significant correlation into very poor visibility conditions (paris38ee) and achieves a greater confidence overall in correlation than the un-enhanced imagery. Ultimately, both enhanced and un-enhanced images run into the noise limit (paris01ee).

C. ALDF Fog Sequence **(see previous gallery of this sequence)**

This sequence of fog clearing at Langley's ALDF was discussed in a previous section. For the purposes of correlation coefficient computation, we reversed the chronological order so the beginning frame is the least turbid image and enhancement and the remaining image/enhancement pairs reflect increasing turbidity. While the images and enhancements were carefully registered spatially, we did note some minimal residual misalignment of one pixel in edge space. Therefore the correlation coefficients for perfect alignment may be slightly higher that the numbers that are given below:

Notes	File Name	R (un-enhanced)	R (enhanced)
fairly clear	fog0213ee	0.29	1.0
increasing turbidity	fog0212ee	0.21	0.60
....	fog0211ee	0.19	0.59
	fog0210ee	0.14	0.56
	fog0209ee	0.16	0.41
	fog0208ee	0.14	0.38
	fog0207ee	0.065*	0.30
	fog0206ee	0.017	0.30
	fog0205ee	0.023	0.24
	fog0204ee	0.090	0.15
	fog0203ee	0.030	0.25
	fog0202ee	0.045	0.21
	fog0201ee	0.029	0.31
	fog0200ee	0.048	0.18

*near detection limit

For these data we see the strongest impact of enhancement on improvement in R values, as well as the strongest maintenance of high correlation values as poor visibility deepens.

We also attempted to do correlation studies on Mars Odyssey time series data but regrettably this proved to be an intractable problem in practical spatial registration. We can only assume that variable orbital parameters, and arbitrary non-nadir viewing combined to make spatial registration a problem where spatial registration varies across the image frames in the time series. As a result, we could not examine correlation for these data, and this is especially disappointing since this data is rich in terrain diversity and we could have learned much more about scene content and context effects of ultimate correlation performance.

VI. D. Experimental Determination of Correlation Coefficient Statistics for Mis-Matched Edge Pattern Frames

In order to gain insight into the issue of what correlation coefficient values can be considered to be a significant "detection" or "frame pattern recognition", we computed R for a wide array of non-matching image edge patterns. Two data sets were specifically aerial images taken from aircraft in flight, and these two sets include a large number of images of airports and runway approaches. Further we computed R for 4 additional data sets on typical terrestrial scenes chosen for maximum diversity of pattern information. For all these cases the mean and standard deviation of R were determined. If we assume that R for mismatching image patterns is a random variable, we can estimate the value of R above which we can have a high confidence level that a legitimate detection of a pattern match has occurred. The data is summarized in the following tables:

For the aerial image patterns:
Data set 1 - 36 image samples
R mean = 0.031 R std dev= 0.029

Data set 2 - 39 image samples
R mean= 0.028 R std dev= 0.015

For the non-aerial images:
Data set 3 - 34 image samples
R mean= 0.005 R std dev= 0.025

Data set 4 - 40 image samples
R mean=0.012 R std dev= 0.022

Data set 5 – 40 image samples
R mean= 0.018 R std dev= 0.017

Data set 6 – 40 image samples
R mean=0.021 R std dev= 0.026

The large scale averages for all data are (weighting all same, and averaging):
R mean= 0.019 R std dev=0.022

The aerial imagery data exhibit a small core pattern similarity that isn't as pronounced in regular imagery, The closer to zero R mean values for regular images is what we would expect for any image data set of unrelated imagery. So the aerial images seem to correlate more even when there is no real pattern match.. To test the validity of the sample size, 3 more sets of 30 images each were used to computer the mean and standard deviation of R. when these were added into the large scale aggregate values, there was no change in the mean and only one digit lowering(least significant) in the standard deviation. From this we conclude that we have probably do have a sufficiently large scale sample to have a reliable set of values for the mean and standard deviation of R for mis-matching patterns.

From the large scale averages of mean and standard deviation for R, we can conclude that:

A mean+2 std dev for a 95% confidence level in a correct match is R>0.063

This value should still be taken as a preliminary estimate whose accuracy should be improved by larger aggregate sampling if edge frame correlation is a part of an ultimate automatic hazard detection scheme. As it stands the number does serve well as a useful baseline estimate which helps to quantify the thinking forward toward defining a future hazard detection system. The fact that this R value is as low as it is seems to stem from the extremely high degree of mathematical uniqueness of the pattern structure extracted from any arbitrary scene with real world complexity.

The value of R>0.05 for significant pattern detection used earlier in the correlation analysis of times series was determined from less extensive data, but is not that far off from this larger scale sample. If the slight bias toward more correlation among aerial pattern images holds up with still larger samples, then a slightly higher R value as a threshold for significant detection of an edge frame would be justified for the aviation in-flight imaging case, and similarly may be slightly lower for aviation imaging during runway taxiing. These R values for significant detection apply just to detecting a match for an entire edge frame. Determination of significance for specific object hazards is not yet defined since specific hazards will be a regional pattern within the whole edge frame and would not be expected to have the same high degree of pattern uniqueness as an entire scene's extracted edge pattern.

VII. Conclusions

The extensive galleries of time series image enhancements and their associated edge patterns provide a very broad comprehensive set of data to support the idea that the RVS processing is a “universal front-end” for approaching canonical visual representation with real-time processing and enables a high degree of pattern constancy. This serves as a robust starting point for higher level pattern recognition and the practical implementation of vision-based automation in aviation safety applications, specifically smart imaging sensors with built-in real-time external hazard detection. In addition, a preliminary correlation analysis was presented which quantifies the degree of pattern constancy and the impact of RVS image enhancement processing on achieving resilient pattern recognition for aviation sensors of the future. Together these data cover wide ranging cases of extraneous imaging variables that represent a major obstacle to computer pattern recognition and demonstrate a very significant diminution of these obstacles in a very general sense which makes practical generic computer visual pattern recognition a realistic technology goal for the immediate future.