First RHEA International Conference on Robotics and associated High-technologies and Equipment for Agriculture
Hosted by the University of Pisa
Pisa, Italy, September 19-20-21, 2012

TOPIC n° 2.2

Real-time weed/crop discrimination through fast direct image registration

Xavier P. Burgos-Artizzu[1], Piotr Dollár[2], Angela Ribeiro[3], Pietro Perona[1]

[1] *Computation and Neural Systems, MC 136-93.California Institute of Technology. Pasadena, CA, USA (xpburgos@caltech.edu, perona@caltech.edu)*

[2] *Microsoft Research, Redmond, WA, USA (pdollar@microsoft.com)*

[3] *Center for Automation and Robotics (CAR).UPM-CSIC. Arganda del Rey, Madrid, Spain. (angela.ribeiro@csic.es)*

Keywords: Weed/crop discrimination, Crop row detection, Real-time image processing, Video stabilization, Automatic vehicle guidance

---

Abstract: This paper presents a computer vision system that is able to discriminate between weed patches and crop rows in real-time, from videos taken directly from a tractor moving through the field. Weed/crop discrimination is highly simplified thanks to video stabilization. We present a simple but effective variant of the inverse compositional algorithm for image alignment, and show that on our videos, our optimized version of the algorithm performs just as well as key-point matching methods, while being up to 2x faster. Once the video stabilized, crop rows remain almost constant through short periods of time, and be detected by a simple image processing. We tested our approach on several videos, taken in different maize fields on different dates, and presenting a variety of weed/crop conditions. Our final approach achieves a mean recognition of 84% on weeds and 91% on crop pixels, improving on our previous work 9% and 29% respectively.

---

1. Introduction

Real-time weed/crop discrimination is a desired outcome in many applications of precision agriculture. This paper presents a computer vision system that is able to discriminate between weed patches and crop rows in real-time, through prior video stabilization.

We present a simple but effective variant of the inverse compositional algorithm for image alignment introduced in (Baker and Matthews, 2001, 2004). This direct method has fallen out of use, in favor of interest point matching methods, such as the widely used Scale-Invariant-Feature-Transform, SIFT, (Lowe 1999). Although it is true that SIFT outperforms direct registration methods on standard, rich images, it does not so on less structured images, and especially in continuous video. Homogeneous images such as those of a crop field make the correct matching of interest points difficult. Also, in a video running at 25 frames per second, the change across frames is small, and therefore direct registration methods work well. We show that on our videos, our optimized version of the inverse compositional algorithm performs just as well as SIFT, while being up to 2x faster, running at 40 fps in a single core CPU.

Once stabilized, crop rows remain almost constant through short periods of time, and weeds can therefore be discriminated by their position between crop rows and their movement across frames by a simple image processing.

2. Materials and methods

Section 2.1 introduces the proposed variant to the Inverse Compositional Algorithm for image alignment, while Section 2.2 outlines the image processing for weed/crop discrimination.

2.1 Inverse Compositional Algorithm

The first step of the proposed method is to stabilize the video. Each frame is registered to its next using an optimized version of the inverse compositional algorithm for image alignment, introduced in (Baker and Matthews, 2001).

The input to the alignment algorithm is a reference image $I_0$and a transformed version of I_0, denoted by I. The goal is to recover a transform $T \circ I = I_0$,where $T \circ I$ denotes application of the transform T to the image I. We assume that the transform T comes from a known set of continuous transforms with k degrees of freedom. We also use $(T_0 \circ T_1)$ to denote an operator on two transforms.

Let T^{δ} denote δ applications of T. The key behind direct approaches to estimating a transform T is the assumption that for small δ, the following holds:

$$T^{\delta} \circ I \approx I + dI \cdot \delta \tag{1}$$

Here, dI is the first order approximation of $T^{\delta} \circ I$, $\mathrm{dI} = (T^{\delta} \circ I) - I$, see Fig. 1. This linearity assumption is directly related to the smoothness of the manifold, which in turn is related to the smoothness of the image itself. Smoothing images prior to applying the method makes (1) more accurate. However, smoothing an image results in a loss of information, so the amount of smoothing has to be chosen carefully.

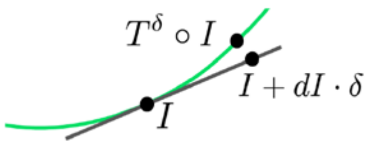

Fig. 1. Direct methods for image alignment assume that applying a small amount δ of a transform T can be represented reasonably well by a first (or second) order approximation.δcan be fractional.

Let $T_1 \cdots T_k$ represent a set of k basis transforms, such that any transform can be written as $T = T_1^{\delta_1} \circ T_2^{\delta_2} \circ \cdots \circ T_k^{\delta_k}$. Applying equation (1) and dropping higher order terms, we get:

$$T \circ I \approx I + \nabla I \delta \tag{2}$$

where ∇I is a $n\ x\ k$ matrix with each column k set to $(T_k \circ I) - I$. Given (2) and combining it with (1) we finally get:

$$I + \nabla I \delta \approx I_0 \tag{3}$$

See Fig. 2 for a visualization of equation (3).We can solve for δ using least squares. To ensure the estimation is well conditioned we perform Tikhonov regularization, encouraging small $\|\delta\|_2^2$:

$$\delta = (\nabla \mathrm{I}^{\mathrm{T}} \nabla \mathrm{I} + \lambda \mathrm{E_k})^{-1} \nabla \mathrm{I}^{\mathrm{T}} (I_0 - I) \tag{4}$$

Here $\mathrm{E_k}$ is the $k\ x\ k$ identity matrix and λ is a small constant set to 10^{-6} in all reported experiments. Finally, given δ, we can compute $T = T_1^{\delta_1} \circ T_2^{\delta_2} \circ \cdots \circ T_k^{\delta_k}$.

The procedure described above can be improved further by warping I according to the recovered transform T to yield

$I' = T \circ I$, then solving with I' in place of I. This gives a new transform such that $T' \circ I' = T' \circ T \circ I \approx I_0$.

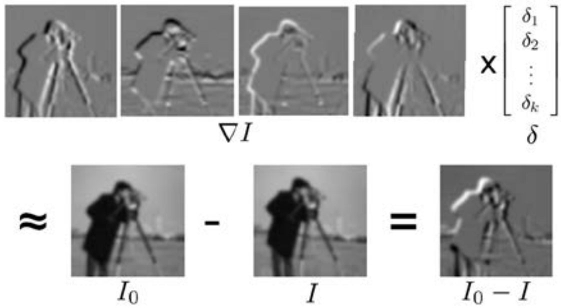

Fig. 2. A visual demonstration of equation (3).The derivative images ∇I, combined using δ, give rise to the difference $I_0 - I$. Given ∇I, I_0 and I, least squares can be used to compute δ.

The recovered transform after two steps is the *composition* $T' \circ T$ and this procedure can be iterated until convergence (in practice only a few iterations are necessary). Note that in each iteration we must recompute ∇I, which can be fairly costly. To avoid this computation, we can reverse the role of $I_0$and I, solving for a transformation $\mathrm{T_0} \circ \mathrm{I_0} = \mathrm{I}$, and then applying the *inverse* transform T_0^{-1} to I at each iteration. The resulting approach is identical except $\nabla \mathrm{I_0}$ needs to be computed only once. As a result, the above algorithm is quite fast and can be implemented in about a dozen lines of code.

2.2 Image processing for weed/crop discrimination

The homographies computed using the Inverse Compositional Algorithm provide information on the displacement that occurred between two frames, and can therefore be used to align frames on top of each other. By doing this, all jumps in the image coming from tractor jolting or sudden lateral displacements are smoothed. Since the tractor travels parallel to the crop rows, the result is that in a stabilized video crop rows position is more stable, see top row of Fig. 3. Therefore, after vegetation is segmented from images (using the same approach as in Burgos-Artizzu et. al., 2011), crop rows can be detected with a simple AND operation over time, leaving weeds as the remaining vegetation pixels (after cleaning the image).

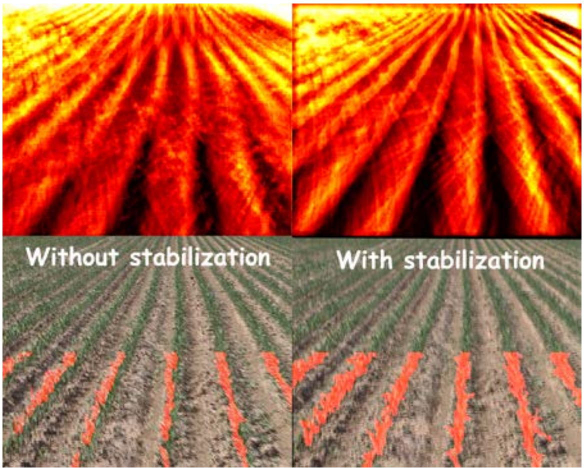

Fig. 3. Crop row detection comparison with video stabilization vs. no stabilization on movement video sequence. Top row: accumulated image over the first 3 seconds of a video. Bottom row: result of AND operation between frame t and frames $[\boldsymbol{t-1}...\boldsymbol{t-9}]$.

3. Results and conclusions

Fig. 4 shows alignment results on frames from movement sequence, comparing the Inverse Compositional Algorithm (Homog) and a SIFT based image alignment. Homog is twice as fast, while performing similarly. Both methods can be applied at full image resolution or at smaller scales, for a trade-off between precision and speed. We use Homog at 1/4 resolution,
which shows an alignment error only 4% superior to that of full resolution, while running 13x faster (40fps).

Table 1 shows weed/crop detection results on the same video sequences used in (Burgos-Artizzu et. al. 2011). The new method outperforms previous work in every video sequence, reaching an average 84% correct weed detection and 91% on crop, a 9% and 29% improvement respectively over (Burgos-Artizzu et. al. 2011). The method performs in real-time, at approximately 30 frames per second, on a single core CPU. The method robustness to tractor jolting and terrain irregularities suggests its future possible use for automatic guidance of agricultural vehicles through crop row detection.

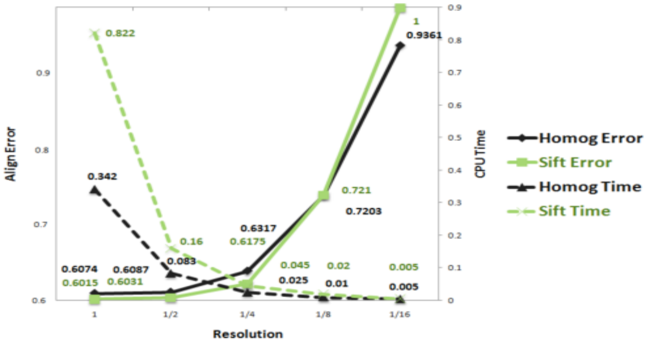

Fig. 4. Error and CPU Time comparison between SIFT and Inverse Compositional Algorithm (Homog) methods on 1500 576x720 frames from 6 different videos. Error is computed as the sum of squared differences between aligned and target image. Both algorithms are run at 4 different image resolutions, to find a trade-off between speed and precision.

Table. 1. Main weed/crop discrimination results on video sequences from (Burgos-Artizzu et. al. 2011)

	Fair2		Sowing Err		Patches		Movement		Average	
	Weed	Crop	Weed	Crop	Weed	Crop	Weed	Crop	Weed	Crop
(Burgos-Artizzu et al. 2011)	93%	83%	64%	65%	65%	70%	74%	36%	75%	62%
Proposed approach	86%	93%	64%	88%	89%	86%	96%	98%	84%	91%
Difference	-7%	+10%	0%	+23%	+24%	+16%	+22%	+62%	+9%	+29%

Acknowledgements

First author holds a postdoctoral scholarship from the Spanish Ministry of Education and Science (MEC), ProgramaNacional de Movilidad de Recursos Humanos del Plan Nacional de I-D+i 2008-2011. The Spanish Ministry of Education and Science (MEC) and the European Union have provided full and continuing support for this research work through projects Plan Nacional AGL2011-30442-C02-02 (GroW) and UE-CP-IP245986-2(RHEA).

References

S. Baker and I. Matthews (2001). *Equivalence and efficiency of image alignment algorithms*. Proceedings of the International Conference in Computer Vision and Pattern Recognition.

S. Baker and I. Matthews (2004). *Lucas-kanade 20 years on: A unifying framework*. International Journal of Computer Vision, 56(3):221-255.

X. P. Burgos-Artizzu, A. Ribeiro, M. Guijarro, and G. Pajares (2011). *Real-time image processing for crop/weed discrimination in maize fields*. Computers and Electronics in Agriculture, 75(2):337-346.

D.G. Lowe (1999). *Object recognition from local scale-invariant features*. Proceedings of the International Conference on Computer Vision.