3D Tracking of Honeybees Enhanced by Environmental Context

Guillaume Chiron[1], Petra Gomez-Krämer[1],
Michel Ménard[1], and Fabrice Requier[2,3]

[1] L3I, University of La Rochelle, Avenue M. Crépeau, 17000 La Rochelle, France
[2] INRA, UE 1255, UE Entomologie, 17700 Surgères, France
[3] CEBC, CNRS, UPR 1934, 79360 Beauvoir sur Niort, France
{guillaume.chiron,petra.gomez,michel.menard}@univ-lr.fr,
fabrice.requier@magneraud.inra.fr

Abstract. This paper summarizes an approach based on stereo vision to recover honeybee trajectories in 3D at the beehive entrance. The 3D advantage offered by stereo vision is crucial to overcome the rough constraints of the application (number of bees, target dynamics and light). Biologists have highlighted the close scale influence of the environment on bees dynamics. We propose to transpose this idea to enhance our tracking process based on Global Nearest Neighbors. Our method normalizes track/observation association costs that are originally not uniformly distributed over the scene. Therefore, the structure of the scene is needed in order to compute relative distances with the targets. The beehive and especially the flight board is the referent environment for bees, so we propose a method to reconstruct the flight board surface from the noisy and incomplete disparity maps provided by the stereo camera.

Keywords: stereo vision, honeybees 3D tracking, surface reconstruction, beehive monitoring.

1 Introduction

Forced to observe the worldwide decline of honeybees (Apis mellifera), biologists began to study different hypotheses that could explain the phenomenon. Recently, the authors of [7] highlighted the evidence of behavioral alterations caused by pesticides. In that study, entrances and exits data was collected by a Radio-frequency identification monitoring device placed at the beehive entrance. So far, no biological study has been conducted at a big scale on flight behaviors. In cause, the lack of suitable methods to collect trajectories of honeybees in flight. The only method used by biologists (harmonic radar) is intrusive and suffers from biases. We believe that computer vision can effectively achieve this task with the respect of the application constraints.

Our challenging application of tracking bees in 3D at the beehive entrance has been studied in [4], laying down the application constraints and a detect-before-track approach based on Kalman Filter coupled with Global Nearest Neighbor. In this current paper, we focus our attention on the weakness of our tracking algorithm nearby the flight board. As a main contribution, we propose to enhance

the tracking process by taking in consideration the targets relative to environment. This is achieved by a normalization of track/observation association costs relying on the distance of the target from the flight board. Also, as a second contribution, we introduce a method to recover the flight board surface, which is not adequately represented by individual disparity images provided by the stereo camera.

This paper is organized as follows. First, Section 2 summarizes the base of our monitoring system, detailing the acquisition method, the segmentation, and the tracking processes. Then, Section 3 introduces in first, our normalization method of track/observation association costs, and then our method to recover and model the flight board surface in 3D. Results are shown and discussed in Section 4. And finally, Section 5 concludes and opens promising perspectives.

1.1   Related Work

In the following are discussed papers dealing with trajectometry based on video related to insects or animals. The process of recovering trajectories can be split into two distinct parts which are presented below: detection and tracking.

For target detection, several methods have been proposed. Detect-before-track approaches are generally based on a segmentation process, and then the observations are associated in order to know tracks using an assignment method. Many methods based on that approach use a more or less advanced background subtraction (e.g. [1,3]). In [3], potential false alarms are filtered using a shape (ellipsis based) matching process. In contrast, some methods introduced do not require any background subtraction. The authors of [13] detect bees using the well-known Viola-Jones method [18], and the authors of [11] introduce an approach based on vector quantization which is able to detect individual bees among hundreds of walking bees. In [16], flying bats are detected taking the advantage of multiple cameras by directly applying Direct Linear Transform. In case of track-before-detect approaches, the position of the target is first estimated, and a probability for that estimation to correspond to the target drives the next estimation. For that kind of approach, a likehood function based on appearance models (pre-computed "eigenbees" in [10], or adaptive models in [17,12]) are used.

Many methods have been proposed for tracking. When following one clearly detected target moving along a simple trajectory, approaches such as those used in [6], near neighbor or mean shift may be sufficient. But in the case of tracking multiple targets, assignment problems arise due to missed detections and occlusions and thus cause false alarms. The authors of [1,3] use Global Nearest Neighbor (GNN) for track assignment, instanciation and destruction. In [16] the authors track multiple flying targets thanks to a multiple hypotheses tracker (MHT). Compared to GNN, MHT integrates the time in the assignment decision process. In [10,17,12,9], a nonlinear motion model is considered for their targets and they base their tracking on a particle filter [15], which corresponds to a track-before-detect approach. In [9], the authors introduced a MRF-augmented particle filter for multiple targets with reasonable computational cost,

contrasting with the three other methods used in [10,17,12] that are less suitable in term of performance when working on multiple interacting targets.

When dealing with many flying targets in natural conditions (uncontrolled light and background), the range of application remains narrow and methods remain to be explored. Compared to most approaches mentioned above, our stereo vision based system brings in this paper the advantage of the the 3rd dimension. This work extends the preliminary study made in [4] by considering the spatial environment into the tracking process.

2   Detect-before-Track Based on Stereo Vision

This section presents our stereo vision acquisition system and summarizes how we recover trajectories from a sequence of couples of intensity/disparity images. More details on the acquisition system and our segmentation are given in [4].

2.1   Acquisition System and Constrains

The following constraints make our application especially challenging: Up to 15 simultaneous flying targets with chaotic dynamics, uncontrolled lighting, and gradual soiling of the background. Among potential suitable cameras (stereo vision and time of flight), we chose the TYZX G3 EV stereo vision camera which seemed to satisfy the best our requirements, with the following configuration: 3 cm baseline and 62 degree HFOV lenses. The camera targets the flight board and is located 50 cm above it. This system acquires a couple of intensity/disparity images of 752×480 px at an average frame rate of 47 fps (unstable rate). Figure 1 shows a capture sample. The quality of the disparity map provided by the camera depends on lighting and texturing conditions. Bees standing on the flight board measure about 10 pixels on those images, and during high activity periods, the flight board can be crowded of bees, making the board difficult to discern.

Fig. 1. Couple of intensity (left) / disparity (right) images provided by the stereo vision camera. The beehive is visible on the top. Bees take off/land on the flight board.

2.2   Segmentation of Flying Targets

Under the constraints listed above, classical motion detection methods based on background modeling fail. Our system detects flying bees thanks to an hybrid segmentation that takes advantage of both intensity and disparity images provided by the stereo camera. Our segmentation relies on an adaptive background

model for the intensity, and combined mathematical morphologies and adapted thresholding for the disparity images. As a result, we obtain a mask containing regions corresponding to our targets. A target is defined by (u, v) the center of mass of the region, and d the median value of the depth values in the disparity image. So far, targets are expressed in the image reference coordinates. To ensure later a coherent tracking, targets are rather defined by (x, y, z) in the camera coordinate space (3D Euclidian space) using a projection based on stereo-camera calibration parameters. Our segmentation is globally robust but still returns in rough conditions up to three false alarms per image around the flight board. Also relatively few miss detections occur for too fast, too close or too far targets from the camera.

2.3 Multi-target Tracking in 3D

In our detect-before-track approach, each target is tracked by a Kalman Filter [8]. Despite the apparent rough dynamic of bees, we acquire frames at a sufficiently high frequency (about 47 fps) so that we can assume a constant speed model. Noise matrices (Q for the model, R for the measure) are tuned to prevent the model from derivation. Let us suppose $Y_{1:n}$ the series of observations corresponding to a target from time 1 to n. A Kalman Filter is instantiated with Y_1 and later destroyed when the step $k > n$. For a given step k, an observation is defined by the vector $Y_k = [x, y, z]^T$, and the estimated state of a target is defined by the vector $X_k = [x, y, z, \dot{x}, \dot{y}, \dot{z}]^T$ combining its 3D position and velocity. Figure 2 lays down the recursive mechanism of the Kalman Filter for the estimation of the state vector X_k and the prediction X_{k+1} given an sequence of observation from step 1 to k.
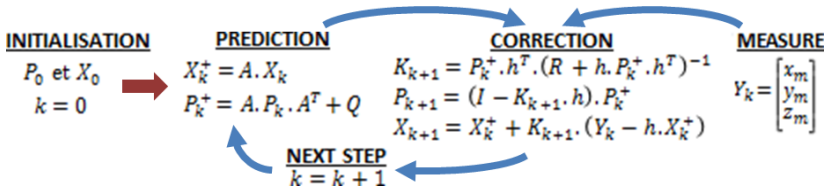

Fig. 2. Recursive mechanism of the Kalman Filter for the estimation of X

In our application, up to 15 targets can be observed at each step, which requires a multi-target tracking approach. The commonly used Global Nearest Neighbor (GNN) [2] method handles observation to track associations, track instantiations and destructions. In GNN, the assignment matrix $A[c_{i,j}]$ represents all the possible associations and the costs generated by those associations. A includes the possibility for each observation to be associated to an existing track, not to be associated to any track or to be associated to a new track. $c_{i,j}$ is the cost for the observation i to be assigned to the possibility j. The best configuration of associations is the solution that minimizes the global cost (e.g. solved

by the hungarian method). The association cost between an observation and a track is computed by the Mahalanobis distance given by:

$$d^2 = (Y - MX^+)'S^{-1}(Y - MX^+)$$
$$S = ME^+M' + Em$$
(1)

where Y is the measure vector corresponding to the observation, M the measurement matrix, X^+ the predicted a priori position, S the innovation covariance with Em as the measure noise matrix, and E^+ the predicted noise covariance matrix.

3   Tracking with Environmental Consideration

In GNN, the multi-target assignment process relies on comparisons of association costs d defined by (1). For a track/observation association, if the minimum cost exceeds a fixed threshold later called Association Cost Limitation (ACL), the association is not considered and the observation becomes a false alarm or a new track. The global chaotic dynamics of flying bees forces the adoption of a tolerant threshold. In our application the problem is, in addition to an high frequentation, most of wrong detections (false alarms and miss detections) are located near the flight board. Moreover, targets near the flight board tend to progress nearly on the same plan resulting in the lost of the 3D advantage. So when tracking a bee near the flight board, in case of miss detection the loose threshold allows the track to be associated with a nearby false alarm or other candidate observations, which causes the degeneration of the track. The magnitude of association costs is not uniformly spread over the scene, so adapting the association costs to the situation is needed.

3.1   Association Costs Normalization

As reference, a classic approach consists in adapting the uncertainty at the track initialization. Bees entering the scene from the outside generate tracks with a big initial uncertainty, which has the effect of decreasing d defined by (1). In contrast, bees taking-off from the flight board have a low velocity and thus can generate tracks with a small initial uncertainty, which has the effect of increasing d. In this way, tracks are less vulnerable to wrong associations nearby the flight board and more able to be associated to relatively far observations when entering the scene. The problem is only partially solved because, the convergence of the uncertainty (in the Kalman Filter) is naively driven by the elapsed time, which does not corresponds well to the random chaotic trajectories exhibited by honeybees.

The authors of [14] explain the relation between honeybees' speed and surrounding objects. Using our classic tracker based on [4], we effectively observed lower speeds and also lower association costs nearby the flight board. The lower association costs are explained by the more stable targets' dynamic resulting in easily predictable positions. As a contribution, we propose to normalize association costs all over the scene relying on targets' relative distances from the board. Figure 7 shows an example of association cost distribution, within which

we can identify the two following distinct effects. The primary growing effect
($l < 50$) potentially corresponds to the gradual stability in the dynamic adopted
by bees when approaching the flight board in the last centimeters. The second
effect ($l > 50$) is less perceptible, and potentially corresponds to changes in bees'
dynamic at a larger scale (e.g. approach of the beehive). Relying on the modeling
(functions α and β) of those effects, we propose the normalized cost c_f based on
d^2 that has the advantage to penalize potential wrong detections near the board
and thus limits the degeneration of the track. c_f is given by:

$$c_f = \begin{Bmatrix} d^2/\alpha_f & \text{if } f \le l \\ d^2/\beta_{f-l} & \text{if } f > l \end{Bmatrix} \tag{2}$$

where f is the closest distance of the target from the flight board, l the limit of
strong influence of the environment, and α and β are respectively the quadratic
and affine normalization functions of f relative to the application. Coefficients
of α and β can be estimated by quadratic and linear regressions from a set of
association costs relative to the board distance taken from well recovered tracks.

The relative distance target/flight board can be computed only with the
knowledge of the structure of the flight board surface. Therefore, the follow-
ing section proposes a method to reconstruct the flight board surface under the
constraints of our application.

3.2   Surface Reconstruction in Cluttered Conditions

Individual disparity images provided by the stereo camera are noisy, incomplete
and do not represent the flight board surface as needed. Therefore, we need to
compute from a sequence of disparity images a sample of pixels that represent
the surface as complete as possible and without any bee on it. As a paradox, a
medium activity constitutes an optimal condition to retrieve the structure of the
flight board. Indeed, zero activity means that untextured parts remain untex-
tured over all the sequence, which is not good for the disparity computation. In
contrast, an overcrowded flight board makes it hard to distinguish the surface. In
the following we propose a method that iteratively filters pixels from a sequence
of disparity images that represent the best the flight board structure.

A region of interest corresponding to the flight board is defined manually, and
the flight board is isolated from the rest of the image. A median filter is applied
on each disparity image in order to filter inconsistent disparities due to stereo
matching errors. A model is initialized with the first image of the sequence. Then
image after image, we update each pixel of the model with the greater disparity
value between the model and the current image. Figure 3 shows step by step the
flight board being extracted. N is the number of disparity images used. It has
to be high enough to include a sample of complementary depth maps that could
be robust to the clutters by passing through it. But, due to the maximization of
the disparity, an overestimation of N increases the chances to add noise (wrong
disparity values that have not been filtered by the median filter). An optimal
number of iterations (or number of disparity images to use) can be estimated
by finding the minimum standard deviation of the depth values on the board

over iterations. Figure 5 shows an example with 180 as the optimal number of iterations which corresponds to a flight board approximately located at 380 mm from the camera.
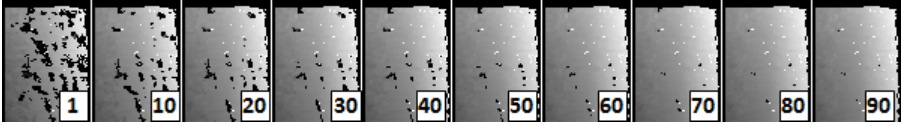

Fig. 3. Step by step flight board reconstruction (90° rotated, partial top view). The numbers are the number of disparity images used. The darker the closer from the camera. Black areas correspond to unavailable disparity information.

As the model still contains holes and inconsistencies, we then apply a fitting based on the locally weighted scatterplot smoothing method (LOWESS) regression modeling method [5] with the Tukey's bisquared function for outliers resistance. Figure 4 shows a mesh reconstructed from the surface model representing a curved flight board.
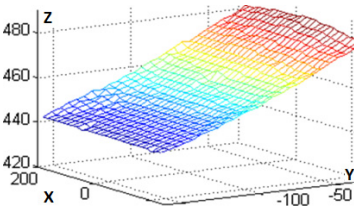

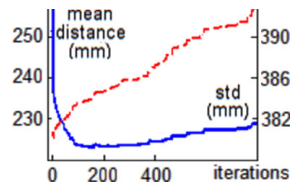**Fig. 4.** Reconstructed surface model representing a curved flight board

Fig. 5. Standard deviation (blue line) and mean values (red dashed line) of depth values of the flight board under reconstruction

4   Results and Discussion

Since no trajectory ground truth is available for our application, we based our evaluation on simulated bee trajectories using a multi-agent approach. Our simulator is based on the following assumptions: Bees tend to slow down nearby the flight board and also when changing direction. Bees located at the limit of the field of view can reach a maximum speed of 3 cm/step. We simulated 200 trajectories (100 landings, 100 takeoffs) with a constant number of 15 bees in the field of view. At each step of the simulation (or frame), 3 false alarms are added, and each bee has 10% changes to be undetected. Those conditions of simulation result in similar data obtained from the terrain under challenging conditions. Figure 6 shows a set of simulated/recovered trajectories with respect to a real reconstructed flight board. The shape of the trajectoires are unique but still tend to follow a similar pattern.
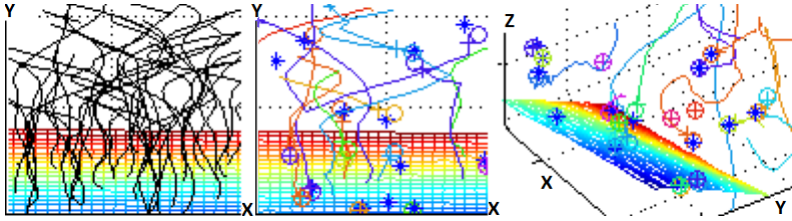
Fig. 6. Top view of simulated trajectories (left) based on a real recovered flight board, and simulated bees tracked with our normalized method (center). 3D view of bees tracked from a real capture with our normalized method (right). Crosses and circles respectively correspond to observations and predicted targets position.

Figure 7 illustrates the effect of our normalization on the relation association cost / distance from the flight board. The normalization model is tuned as followed: Area A with a polynomial function having as coefficients 0.00048, -0.0029 and 0.038 and Area B with an affine function having as coefficients 0.0013 and 1. Therefore, normalized costs can be compared to a constant ACL.
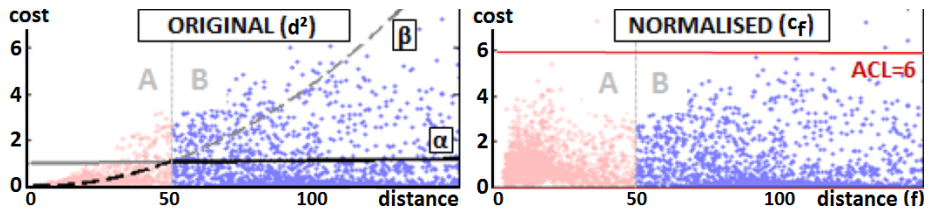

Fig. 7. Association cost / distance to the flight board relation between the original (unnormalized) and our normalized method. The lower original costs nearby the board are explained by easily predictable moves. Areas A and B are separated at the distance of $l=50$ mm. The quadratic function α (dashed curve) and the affine function β (plain line) are obtained respectively by fitting of points of A and B. As an example, ACL=6.

Figure 8 shows comparative results between the original and our normalized method according to different ACL. The classic method recovers at most 60% of the tracks (ACL=7), where our normalized method recovers 71% (ACL=13). Figure 9 shows the importance of recovering accurately the flight board when using our normalized method. Considering an optimal ACL of 13, the result decreases when adding a positive or negative error to the depth of the flight board surface.

If the flight board surface can not be well recovered (e.g. too much activity), it is preferable to manually lower the flight board estimation (positive depth error in Figure 9) to take in consideration the layer of walking bees. This layer measures up to 15 mm, which does not affect results drastically. Without automatic feedback on the quality of the recovered flight board surface, it would be acceptable to consider by default the 15 mm layer of walking bees (being there or not).

The use of a simulation to evaluate tracking results is debatable. Despite our efforts to create a truthful simulator, there will always be differences between simulated and real data. Nevertheless, we confirmed (without quantification) the benefit of our normalization method on real trajectories (see Figure 6). Further studies involving real data and comparisons of simulation methods have already been started.
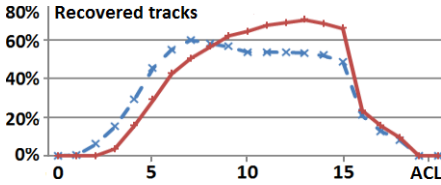

Fig. 8. Comparative results between the unnormalised (dashed line) and our normalized (plain line) methods using different Association Cost Limitation (ACL)
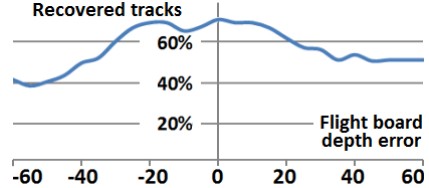
Fig. 9. Effect of a potential wrong flight board recovery (delta depth) on the normalized method based on the optimal ACL, which is 13

5   Conclusion

In addition to our global tracking system based on stereo vision partially presented in [4], we introduced in this article the idea of taking into account targets' environment into the tracking process. The accurate reconstruction of the flight board surface made possible the computation of the relative distance of the bee from its landing/takeoff area. Then our normalization method for association costs made in case of miss detections the tracking process less vulnerable to wrong associations coming from false alarms. Our normalization method provides an improvement of 17.5% on simulated trajectories, which shows the importance taking in account the difference magnitude of association costs when the application allows it. As an example, it works well with our honeybees which tend to follow a typical behavior nearby the flight board.

Comparing costs is a common step in other tracking methods such as MHT, it would be interesting to check the portability of our method. Concerning the flight board reconstruction, our method introduced in Section 3 works with a camera located above the flight board. It could be also interesting to study the possibility of recovering the flight board surface from a non-vertical view.

Concerning long term perspectives, biologists are interested in high level applications such as abnormal behavior detection. On the one hand, behavior models could focus on individual bee trajectories. We can imaging a tracker that takes in consideration more parameters (e.g. activity, weather, time of the day) to adapt motion models. On the other hand, more global models could focus on the colony activity. We can imagine abnormal colony behavior detector based on some basic rules (e.g. low activity during a sunny day).

References

1. Balch, T., Khan, Z., Veloso, M.: Automatically tracking and analyzing the behavior of live insect colonies. In: 5th International Conference on Autonomous Agents, vol. 2001, pp. 521–528. ACM (2001)
2. Blackman, S., Popoli, R.: Design and Analysis of Modern Tracking Systems. Artech House Radar Library, Artech House (1999)
3. Campbell, J., Mummert, L., Sukthankar, R.: Video monitoring of honey bee colonies at the hive entrance. In: VAIB 2008, vol. 8, pp. 1–4 (2008)
4. Chiron, G., Gomez-Krämer, P., Ménard, M.: Outdoor 3d acquisition system for small and fast targets. application to honeybee monitoring at the beehive entrance. In: GEODIFF 2013, pp. 10–19 (2013)
5. Cleveland, W.S.: LOWESS: A Program for Smoothing Scatterplots by Robust Locally Weighted Regression. The American Statistician 35(1), 54–54 (1981)
6. Hendriks, C., Yu, Z., Lecocq, A., Bakker, T., Locke, B., Terenius, O.: Identifying all individuals in a honeybee hive - progress towards mapping all social interactions. In: VAIB 2012 (2012)
7. Henry, M., Beguin, M., Requier, F., Rollin, O., Odoux, J.F., Aupinel, P., Aptel, J., Tchamitchian, S., Decourtye, A.: A common pesticide decreases foraging success and survival in honey bees. Science 336(6079), 348–350 (2012)
8. Kalman, R.E., et al.: A new approach to linear filtering and prediction problems. Journal of Basic Engineering 82(1), 35–45 (1960)
9. Khan, Z., Balch, T., Dellaert, F.: Efficient particle filter-based tracking of multiple interacting targets using an mrf-based motion model. In: IEEE/RSJ International Conference on Intelligent Robots and Systems, vol. 1, pp. 254–259. IEEE (2003)
10. Khan, Z., Balch, T., Dellaert, F.: A rao-blackwellized particle filter for eigentracking. In: Computer Vision and Pattern Recognition, vol. 2, pp. II–980 (2004)
11. Kimura, T., Ohashi, M., Okada, R., et al.: A new approach for the simultaneous tracking of multiple honeybees for analysis of hive behavior. Apidologie 42(5), 607–617 (2011)
12. Maitra, P., Schneider, S., Shin, M.: Robust bee tracking with adaptive appearance template and geometry-constrained resampling. In: 2009 Workshop on Applications of Computer Vision (WACV), pp. 1–6. IEEE (2009)
13. Miranda, B., Salas, J., Vera, P.: Bumblebees detection and tracking. In: VAIB 2012 (2012)
14. Portelli, G., Ruffier, F., Roubieu, F.L., et al.: Honeybees' speed depends on dorsal as well as lateral, ventral and frontal optic flows. PLoS ONE 6(5), e19486 (2011)
15. Ristič, B., Arulampalam, S., Gordon, N.: Beyond the Kalman Filter: Particle Filters for Tracking Applications. Artech House Radar Library, Artech House (2004)
16. Theriault, D., Wu, Z., Hristov, N., Swartz, S., Breuer, K., Kunz, T., Betke, M.: Reconstruction and analysis of 3d trajectories of brazilian free-tailed bats in flight. Tech. rep., CS Department, Boston University (2010)
17. Veeraraghavan, A., Chellappa, R., Srinivasan, M.: Shape-and-behavior encoded tracking of bee dances. IEEE Transactions on PAMI 30(3), 463–476 (2008)
18. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: Computer Vision and Pattern Recognition, vol. 1, pp. I–511 (2001)