I.J.S.N., VOL. 3(1) 2012: 121-124 ISSN 2229 – 6441

INTERNATIONAL JOURNAL OF SCIENCE AND NATURE

www.scienceandnature.org

IDENTIFICATION OF WEEDS IN TURMERIC (*CURCUMA LONGA*) PLANTS USING UNSUPERVISED SKPCA ALGORITHM WITH IMAGE PROCESSING

[1]G.Chinnasamy, [2]M.Palanisamy, [3]M.Pavithra, [4]Azha.Periasamy, [5]Dr.S.Muruganand
[1,3,4,5] Department of Electronics and Instrumentation, Bharathiar University, Coimbatore, Tamilnadu, India
[2]Department of Botany, Bharathiar University, Coimbatore, Tamilnadu, India

ABSTRACT
Turmeric (Curcuma longa) is extensively used as a medicinal plant and food ingredient in India, China and South East Asia. In this paper the implementation of image processing techniques with and Sparse Kernel Principal Component Analysis (SKPCA) is used to identify the weeds in between turmeric growing field. Weeding out the undesirable plants is one of the major activities related to increasing the agricultural production. Herbicide usage can limit the growth of weeds. But the impacts of using herbicide causes the major affects in environment and water contamination, this have led researcher to find the solutions for their accurate use. A machine vision system for weed identification, utilized the morphological properties of weed leaves with stem and pixel value of the leaves. Matching of the image features can be done using unsupervised SKPCA. The accuracy of using this algorithm is 97%.

KEYWORDS: Turmeric (Curcuma longa), Weeds, Image processing, Machine vision, Unsupervised SKPCA

INTRODUCTION
In the present world scenario, India is the world's largest producer (82%) of turmeric (curcuma longa) followed by China (8%), Myanmar (4%), Nigeria (3%) and Bangladesh (3%). Relay 90% of the total production exhibit allelopathy, ploughing and seed purification. Weeds can cause various kinds of losses. In agriculture, weeds can reduce plants yield. Stunted plant growth, will cause the decline. Therefore, the farmers control the weeds, to enhance productivity. A total of fourteen weed species belonging to 8 angiosperm families, were recorded in the fields of turmeric. *Sonchus aspera* L., *Chenopodium album* L., *Rumex dentatus* L., *Ageratum conyzoides* L. *Convolvulus arvensis* L., *Cynodon dactylon* (L.) *Pers., Oxalis corniculata* L., *Malva parviflora* L., *Malvastrum coromandelianum* L., *Trifolium resupinatum* L., *Euphorbia prostarta* L. and *Phalaris minor* Retz, were found to be the most prevalent weed species occurring in 90% or more studied areas during one or the other growing season. The weeds identification and then also deserves attention for effective control. Botanically, weeds species are identification from whole leaf shape and texture venation by weeds specialists, who visually scout fields or study high-resolution images of the field surface. Numerous research activities in the area of plant identification have been initiated in recent years. In the first investigations, different morphological features were extracted from digitised images of plants. These features were compressed and used for identification models (Petry & Kühbauch, 1989). Others extended such models with knowledge-based rules (Guyer *et al.*, 1993). Sökefeld *et al.* (1996) managed to identify 70% of 22 different weed species common in sugar beet by computer vision using morphological features and Fourier descriptors. Rath (1997) presented a plant identification system for identifying native deciduous trees by their foliage leaves. A maximum of 93% computer-based identification of unknown leaves was achieved. This paper deals with the applicability feature matching in image processing using SKPCA (Sparse Kernel Principle Component Analysis) with unsupervised learning algorithm. Destructive impacts of herbicide usage on the environment have led to many researchers orientation toward finding solutions for their accurate use. If density and weeds species could be correctly detected, patch spraying or spot spraying can effectively reduce herbicide usage. A machine vision with precise automatic weed control system could also reduce the usage of chemicals.

MATERIALS AND METHOD
For this study the experimental data of the weed and plant images were obtained from turmeric field. The frequently occurring weeds with absolute frequency of above 80% were *C. album L., M. cromandlianum L.* and *C. dactylon*. Other densely populated weed species with higher absolute density were *A. conyzoides*, *C. arvensis, E. prostrata* and *C. dactylon* [1]. The study highlighted the need to manage weed in order to realize higher turmeric yields. The following figure (1) is the representation of our proposed study

Image acquisition
A Machine vision system included two cameras in the experimental system: a front camera for row guidance and a rear camera for in- row detection of individual plants [6]. The front camera was tilted forward, facing the direction of travel, to allow a 1–2 m section of the plant row to be viewed enabling the vision system to calculate an accurate row center line to position the rear camera directly above the seed-line. The rear camera was mounted directly behind the vision-guided toolbar with the lens pointed vertically down toward the plant row to capture a top view. Determination of image capturing area Image size used was (640x480 dpi with a VGA resolution). Plant

rows were traversed in the same serpentine pattern in which they were planted. Thus the angle of sunlight changed from row to row, field to field, and time to time. The camcorder shutter speed was set to 1:500 of a second to prevent blurred images, but also because outdoor field plants may oscillate at high frequencies due to wind.

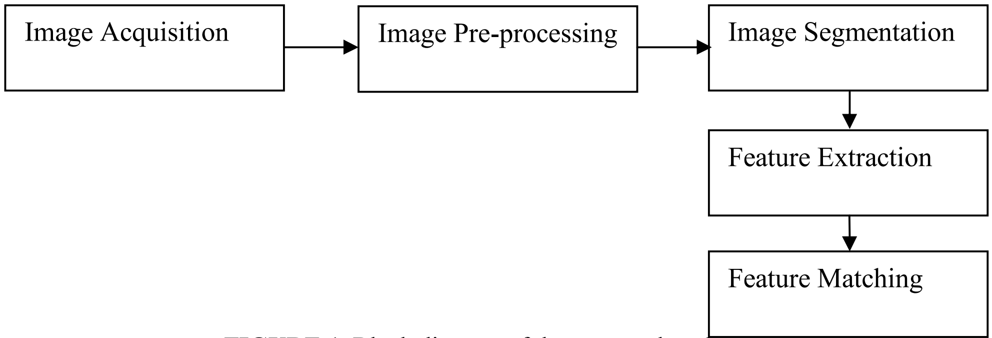

FIGURE 1. Block diagram of the proposed study

Image preprocessing

Turmeric images are acquired from camera in the time interval of 48 hours after 20 days of seeding. Separate the background from the image by using the color transformation. An average 50 images were analyzed per day of containing more than 150 turmeric plants in the field.

Segmentation and object detection

To develop a robust machine vision system for outdoor field conditions, algorithms must be developed to extract useful information in the presence of noise associated with unstructured lighting conditions. Image segmentation is a task of pixel aggregation. After segmentation, each image pixel is assigned into one of several specific classes. Analysis of weed attack distribution is used SKPCA clustering technique.

ALGORITHM DEVELOPMENT

An approach where the frame reduction obeys the output probability maximization criterion was developed, and it is called SKPCA. It consists in estimating the feature space sample covariance for a weed component and the sum of the weighted outer products of the original feature vectors. These weights generate the sparse solution for the KPCA, because they represent a measure of how well a specific training vector contribute to the likelihood maximization. Once obtained the reduced data, the common KPCA technique is applied and the representation of a feature test vector w is given by W_{skpca}. Although the SKPCA generates a reduced training data, it requires the full original training data to evaluate the maximization step, which could be computationally unfeasible, depending on the training data amount. In order to solve it, an approach is proposed, where the original training data is clustered and the SKPCA is applied to these clusters.

$$CF = \sigma^2 I + \sum_{j=1}^{N} \omega_{i} \varphi_i^T$$

Where W is a diagonal matrix composed by the adjustable weights $\omega_1 \ldots \omega_M$, and σ^2 is an isotropic noise component, N $(0, \sigma^2 I)$, common to all dimensions of feature space. It was observed that fixing σ^2, and maximizing the likelihood under the weighting factors ω_i, the estimates of several ω_i are zero, thus realizing a reduced (sparse) representation of the covariance matrix. This approach was based on the probabilistic PCA (PPCA) formulation [5]. Further simplification of this formula is made by the re-estimation of the feature weights and SKPCA of the diagonal matrix can be obtained [6].

$$W_{skpca} = V_F^T \varphi(w) = \Lambda_{K^{-\frac{1}{2}}} \hat{U} T_K k T_w$$

where $\hat{U}_K$ and Λ_K are defined, respectively, as the eigenvectors and Eigen values of $W^{1/2} K W^{1/2}$ and k_w represents the vector calculated by $k(w, x_i)$, where x_i corresponds to the non-zero weighted vectors represented in X. From the figure 3, the feature matching of the weeds and the plants are done using their features. The limits -1 to 0 indicates the weeds and 0 to 1 indicates the plants.

FEATURE MATCHING

1. **Objective approach of the SKPCA with Unsupervised algorithm**
2. The proposed approach consists in making the SKPCA technique computationally feasible for a data set with a great number of samples; the SKPCA re-estimation is computationally unfeasible, once it depends on the kernel matrix K to calculate $\sum$.

3. **Unsupervised Algorithm for training database**

K-means clustering is proposed to divide the full training data into unsupervised clusters algorithm of L frames, and then merge the clusters forming new clusters of 2L frames, which are reduced to L frames by using SKPCA. The process is repeated successively until obtaining just one cluster of L frames, which is the final number of frames desired to represent the full training data. This L frames consists of their own centroids.

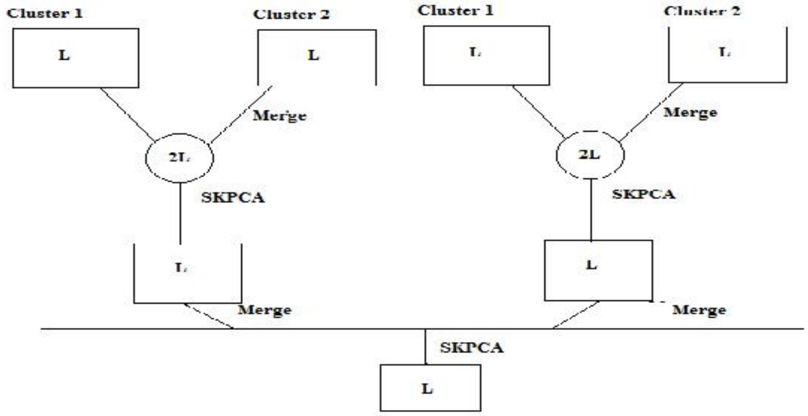

FIGURE 3. Block diagram of SKPCA study

The total number of steps necessary to reduce the l clusters of L frames to just one cluster, is given by step = log2l, where step is the number of steps. Considering this, the proposed approach does not guarantee to reach the overall data maximization, just individual cluster maximization.

4. Maximum Probability

Most relative clusters of the original training data and test samples are matched using their own individual covariance matrix. According to [6], the general form is

$$\mu` = A` \mu b` \qquad \text{...............} (3)$$

The sample covariance matrix is given by,

$$C = 1/M \sum^{M} x_j x^T_j = M^{-1} XX^T \text{...............} (4)$$

Where $X = [x_1 \ldots x_M]$ represents the matrix of data. The M is number of centered observation of the dataset. Both training data and test samples covariance matrix are computed and matched according to their clusters.

RESULTS AND DISCUSSION

Initially image training is done by 50 images of turmeric plants and 60 different images of weeds, the features of these images are stored as a database. The features are extracted from the unknown input images are compared with features already stored in the database. The image acquired is such that it captures both the plant and the weeds in the same frame. Weed segmentation requires two stages. First stage is separation of whole plants from the background. Second stage is separation of weeds from the main plant. Those pixels that are related to plants have greater green pixel value. Before weed plant discrimination pre-processing of the image is done. And applying segmentation to extract the features of the unknown images, these features are matched with the database using SKPCA. The following figure 4a is the occluded turmeric image and figure 4b is the feature extracted image of turmeric plant.

Figure (4a) Occluded Turmeric image

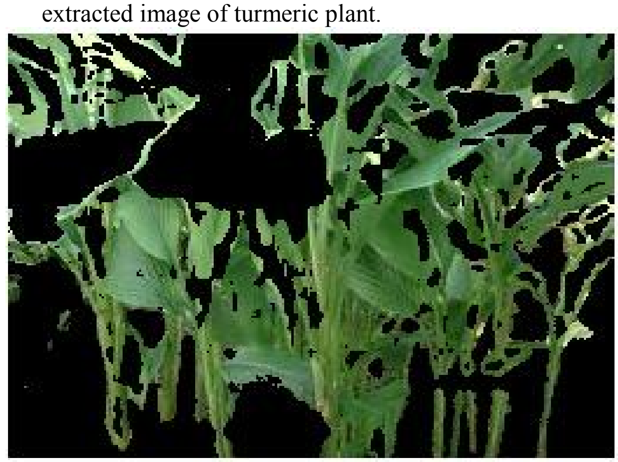

Figure (4b) Extracted Features of Turmeric

In early stage, the plant features were extracted easily. After a certain stage the plant and the weed leaves are overlapped each other the identification becomes more complicated. An important task before feature extraction was the partition and separation of the single plants. From this partition we can easily extract the features; the feature data shows that the pixel value of green in weed has lower influence than the turmeric plant pixel value. So that the identification weeds from the turmeric plant is easily computed. From this below figure (5), the marked region indicates the turmeric plant in the field **[1x1 m^2 (square meter) area]**

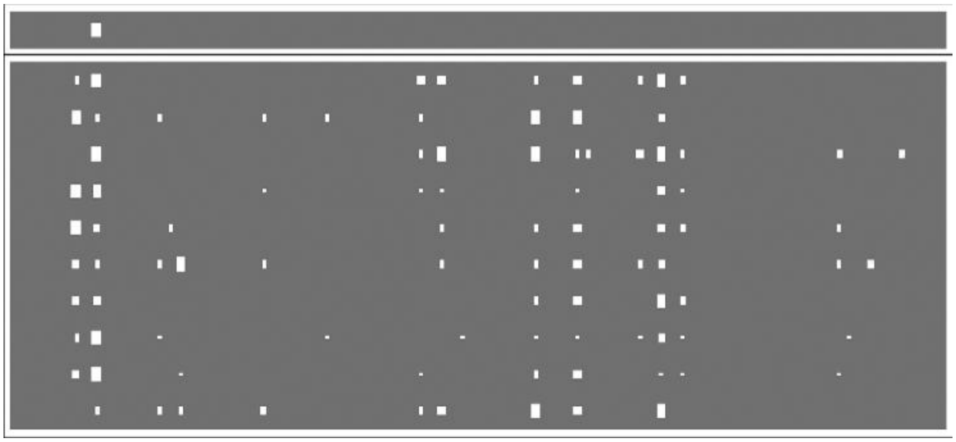

Figure (5) Identified regions of turmeric plants in the field

CONCLUSIONS

In this paper the identification of weed image which has the dominant role in the cultivation of turmeric (Curcuma Longa). A system with high classification of accuracy is established using unsupervised SKPCA clustering algorithm. This algorithm were able to identify the weed density and also capable to distinguish turmeric plant and weed. The accuracy of identification system depends upon the pixel values of the plants and weeds, and the clusters produced by the algorithm. By using this algorithm we can differentiate the weeds and plants at the accuracy rate of 97%. According to this statistics of results the unsupervised SKPCA is very much useful to identify the weeds in turmeric field and very much useful in the yield without weed. In future this algorithm can be used in various plants field to identity the weeds.

ACKNOWLEDGEMENT

We are very grateful to our Department faculty members to provide necessary facilities to do this research work.

REFERENCES

Justina Jane Tahira1, Salik Nawaz Khan, Ruqia Suliman and Waheed Anwar 'Weed flora of *curcuma longa* fields of district kasur, Pakistan' *Pak. J. Weed Sci. Res. 16 (2): 241- 246, 2010.*

Petry W; Kühbauch W (1989). Automatisierte Unterscheidung von Unkrautarten nach Formparametern mit Hilfe der quantitativen Bildanalyse [Automated discrimination of weed species with quantitative image analysis]. Journal of Agronomy & Crop Science 163, 345-351.

Guyer D E; Miles G E, Gaultney L D; Schreiber M M (1993). Application of machine vision to shape analysis in leaf and plant identification. Transactions of the ASAE, 36(1), 163-171

Sökefeld M; Gerhards R; Kühbauch W (1996). Automatische Erkennung von Unkräutern im Keimblattstadium mit digitaler Bildverarbeitung [Automated identification of weeds in the stage of cotyledon using image analysis]. KTBL-Arbeitspapier. In: Innovative Verfahren zur Unkrauterkennung 236, 47-58. Kuratorium für Technik und Bauwesen in der Landwirtschaft e.V. (KTBL). Darmstadt

Rath T (1997). Methoden zur computer- ildanalytischen Pflanzenidentifikation am Beispiel dendrologischer Bestimmungen [Methods for plant identification by image processing demonstrated in the area of dendrological determinations]. Gartenbau technische Informationen, Institut für Technik im Gartenbau und Landwirtschaft. Heft 42**.** University of Hanover, Germany

Lee W S; Slaughter D C; Giles D K (1997). Robotic weed control system for tomatoes using machine vision system and precision chemical application. ASAE Paper No 97-3093

H. Hermansky, "Perceptual linear predictive (PLP) analysis of speech," *Journal of the Acoustical Society of America*, , no. 87, pp. 1738–1752, 1990.

V. Vapnik, *The Nature of Statistical Learning Theory*, Springer- Verlag, 1995.

M.E. Tipping and C.M. Bishop, "Probabilistic principal component analysis," Journal of the Royal Statistical Society, pp.611–622, 1999.

M.E. Tipping, "Sparse kernel principal component analysis," Advances in Neural Information Processing Systems, 2000.

Cheng, H. D., X. H. Jiang, Y. Sun, and Jingli Wang. 2001. Color image segmentation: advances and prospects. *Pattern Identification* 34(12): 2259-2281.

Franz, E., M. R. Gebhardt, and K. B. Unklesbay. 1991. Shape description of completely visible and partially occluded leaves for identifying plants in digital images. *Transactions of ASAE* 34(2): 673- 681.