Automatic Feature Construction and a Simple Rule Induction Algorithm for Skin Detection

Giovani Gomez GEGOMEZ@CAMPUS.MOR.ITESM.MX
Eduardo F. Morales EMORALES@CAMPUS.MOR.ITESM.MX
ITESM – Campus Cuernavaca, Paseo de la Reforma 182 – A, Temixco, Morelos, 62589, México

Abstract

Many vision systems use skin detection as a principal component. Skin detection algorithms, normally evaluate a *single* and thus limited color model, such as HSV, YC_rC_b, YUV, RGB, normalized RGB, etc. Their limited performance, however, suggests that they are looking at the incorrect color models. This paper describes a new constructive induction algorithm that creates adequate attributes for skin detection. The algorithm starts with normalized RGB and is able to produce a single rule with a small number of easy to evaluate terms with superior performance than existing methods. The constructive induction algorithm uses a new restricted covering algorithm, called RCA, as its learning component, which produces a single rule with competitive results when compared against C4.5.

1. Introduction

Skin color detection is an important step in many vision systems, like gesture recognition, hand tracking, video indexing, face detection, etc. (e.g., (Brand et al., 2001; Fleck et al., 1996; Hjelmas & Low, 2001; Kumar & Poggio, 2000; Spors & Rabenstein, 2001; Saber & Telkap, 1998; Sigal et al., 2000; Sobottka & Pitas, 1996; Soriano et al., 2000) to cite just a few). Pixel based skin detection can narrow the search space prior to high-level layers, however this is not an easy task. Skin pixels can vary with ambient light, such as color lamps acting as filters, brightness and specularities, shadows, daylight, etc. Moreover, different cameras return different values for the same scene, hence, skin detection can become a cumbersome task.

There has been a growing interest in using probabilistic methods for pixel based skin detection. One widely-used choice is the Skin Probability Map, or SPM for short (Brand & Mason, 2000; Brand et al., 2001; Jones & Regh, 1999; Zarit et al., 1999), which has been assessed (Brand & Mason, 2000) as the best one in terms of accuracy and running time.

Despite of the long history of skin detection, there are very few works (Brand & Mason, 2000; Zarit et al., 1999; Terrillon et al., 2000) surveying which color models are more reliable for skin detection. Thus, most vision systems just work on HSV, RGB, or normalized RGB, which yields poor results. In this paper, a machine learning approach is used to construct a simple to evaluate skin model with better performance than existing methods.

Conventional inductive learning systems induce models from a fixed set of attributes. For a concept to be learnable, its examples must populate one or more regions of the hypothesis space expressible in the description language. If the original representation produces poor learning results, an alternative approach is to perform automatic transformations of the representation space until a suitable set of attributes is found for the learning task. In this paper we applied a constructive induction approach to find a set of new attributes suitable for skin detection.

Most constructive induction systems use boolean combinations of existing attributes to create new attributes. Their constructive operators can form conjunctions and/or disjunctions of attributes (e.g., (Pagallo, 1990; Ragavan & Rendell, 1993)) or even use more sophisticated operators such as M-of-N (Murphy & Pazzani, 1991) and X-of-N[1] (Zheng, 1995). Although a large number of studies have been devoted to boolean combinations of attributes (e.g., (Zheng, 1998)), there are very few systems that use arithmetic combinations of real-value attributes, which normally occur in vision. Most notably is the Bacon system

[1] M-of-N answers whether at least M of the conditions in the set are true. X-of-N answers how many conditions in this set are true.

(Langley et al., 1983) which searches for empirical laws relating dependent and independent variables. Bacon finds increasing and decreasing monotonic relations between pairs of variables that take on numeric values and calculates the slop relating both terms to create a new attribute. Once a functional relation between variables is found, it is taken as a new dependent variable. This process continues until a complex combination is found relating all the primitive attributes.

In this paper we start with the three basic color components RGB in a normalized form and a simple set of arithmetic operators to produce a suitable model for skin detection. Once a new set of attributes is produced, a new restricted covering algorithm, called RCA, is used to construct a single rule of no more than a small number of easy to evaluate terms with a minimum accuracy. We are interested in inducing simple models as they are relevant to applications which require fast response times, such as, gesture recognition, face and human tracking, etc.

RCA searches for candidate rules in parallel considering two intermixed criteria for selecting new terms, and is shown to produce more understandable and easy to evaluate models, with similar performance, when compared against C4.5.

Section 2 introduces the constructive induction approach followed in this paper. The characteristics of the data used in the test are given in section 3. Section 4 describes the Skin Probability Map (SPM), the current best model for skin detection used in computer vision. The main results of the proposed algorithm are presented in section 5. Finally, section 6 gives conclusions and discusses future research directions.

2. Constructive Induction

The general approach followed in this paper for constructive induction is, as in many other systems, shown in Table 1. The idea, is to start with some primitive attributes and a set of constructive operators, create a new representation space, run an inductive learning algorithm, and select the best attributes of this new space. This process continues until a predefined stopping criterion.

This general approach is composed, as most constructive induction systems, of three basic components working together: (i) the machine learning algorithm, (ii) the constructive induction module, and (iii) an evaluation component.

Our main goal is to induce a single rule with a small number of relatively simple terms, which can be easily evaluated and effectively used for human tracking. We want it also to be competitive with existing skin recognition methods.

Table 1. General constructive induction algorithm.

CurrentAttrib = original attributes
Operators = set of constructive operators
UNTIL termination criterion
- NewAttrib = CurrentAttrib ∪ new attributes constructed with Operators on CurrentAttrib
- Run a machine learning algorithm on NewAttrib
- CurrentAttrib = Select the best attributes from NewAttrib

Our learning algorithm, called RCA (Restricted Covering Algorithm) follows closely a general covering algorithm, with some restrictions. Rather than trying to build a set of rules, RCA tries to build a single rule for each class with no more than a predetermined number of terms. In our case, we are interested in learning just one rule for skin detection.

The general strategy of RCA is to favor attributes which cover a large number of true positives and attributes with small number of false positives. Since we are interested in a single rule we will talk about the total number of true positives (TTP) which will be used to increase the measure of recall[2] and the total number of false positives (TFP) which will be used to increase precision[3].

Since we are dealing with real-value attributes, RCA creates binary splits using an information gain heuristic (the same used by C4.5 (Quinlan, 1993)). RCA considers two possible attributes in parallel when constructing rules. On its first cycle, RCA constructs two rules which have as LHS the attribute with larger TTP in one rule and the attribute with larger $TTP - TFP^2$ in the other rule. The following cycle produces two rules out of each original rule (4 in total) following the same criterion, again adding to the LHS of each rule one attribute with large coverage and one which is heavily penalized by the number of misclassifications. This process continues until the rules produced have a certain number of predetermined terms. The total number of alternative rules produced is 2^n, where n is the number of terms on each rule. RCA builds 2^n rules in parallel aiming for a large coverage with small errors on the same example set. After the termination

[2] Recall $= \frac{TP}{TP+FN} \times 100\%$, where TP = true positives and FN = false negatives.

[3] Precision $= \frac{TP}{TP+FP} \times 100\%$, where FP = false positives.

Table 2. Overall description of RCA

For each class C
  Let E = training examples
  Let N = maximum number of terms per rule
  Create a rule $R(0)$ with empty LHS and class C
  Let depth $D = 1$
  Until $D = N$ **do**
    For each attribute A create a split (Sp_A)
      with greater information gain
    For each existing rule $R(D-1)$
      create two new rules ($R_1(D)$ and $R_2(D)$) by
      adding to its LHS, a Sp_i with larger TTP
      ($R_1(D)$) and a Sp_j with larger $TTP - TFP^2$
      (R_2)
    Let $D \rightarrow D + 1$
    For each $R_i(D)$ continue with its own
      covered examples from E
  Output all $R_i(D)$

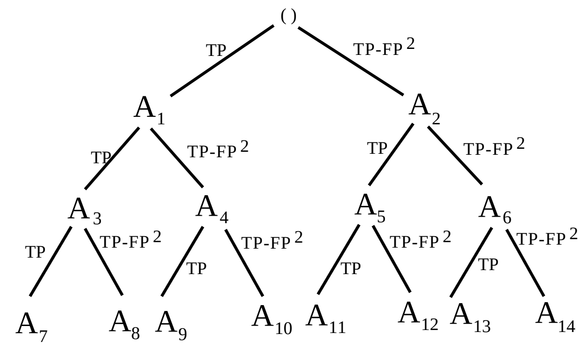

Figure 1. Rules created by RCA. Each branch adds an attribute using either criterion. Each path, from the root to a leaf, represents a rule.

criterion all the alternative rules are displayed along with measures of recall, precision, and success rate. An overall description of RCA is given in table 2. The idea behind constructing rules with two intermixed criteria is to induce rules with complementary attributes. We asses this strategy in section 5 when we compared the performance of RCA with C4.5.

For instance, if RCA is restricted to create rules of three terms, this produces 8 rules in parallel with at most 14 attributes (if there are no duplicates). Figure 1 shows the different rules constructed with three terms. Each path, from the root to a leaf, represents one possible rule defined as the conjuction of the three attributes along the path.

The constructive induction algorithm started its representation space with the three basic color features: RGB in a normalized form (i.e., $\frac{R}{R+G+B}$, $\frac{G}{R+G+B}$, and $\frac{B}{R+G+B}$) and the constant 1/3 which we thought to be useful since we are using three normalized attributes. All of them were used to create new attributes by seven constructive operators: $A+B, A*B, A-B, B-A, A/B, B/A,$ and A^2, where A and B can be any pair of distinct attributes or 1/3 (except for A^2 where 1/3 is not considered).

3. Datasets

We explored the performance of our constructive induction algorithm and its learning component, RCA, in two types of images: (i) skin/non-skin indoor and (ii) skin/non-skin outdoor. We used several daylight and illumination conditions, as well as a number of input sources. These image sets cover more than 2,000 people from different races and ages, and cover a wide range of illumination conditions, from Tungsten lamps (~3200K) to daylight (D_{65} or 5000K-5500K). We expected a shift to yellow in indoor scenes (Tungsten), and a little to blue in outdoor conditions (sun-shine). In the case of outdoor scenes, we use both direct sunshines and shadows. The complete dataset has more than 32 million pixels *carefully* labeled, and it is publicly available by contacting the first author.

SPM was trained on two thirds of the 32 million pixels and tested on one third of them following the recommendations given in (Brand & Mason, 2000). On the other hand, RCA and C4.5 were both trained on 150,000 randomly selected points, and tested on the same subset of 22 million pixels. This subset has a balanced number (11 million) of skin and non skin points.

4. Skin Probability Maps

There has been a growing interest in using probabilistic methods for pixel based skin detection. One widely-used choice is the Skin Probability Map, or SPM for short (Brand & Mason, 2000; Brand et al., 2001; Jones & Regh, 1999; Zarit et al., 1999), which has been assessed (Brand & Mason, 2000) as the best one in terms of recall and precision. A popular version of SPM (Brand & Mason, 2000; Brand et al., 2001; Jones & Regh, 1999; Zarit et al., 1999) is a lookup table where RGB values directly address a voting slot. Two 3D histograms are computed, one for skin and one for non-skin. After dividing every slot by the total count of elements, we get an associated probability on a $[R, G, B]$ index. Statistics are derived from these histograms. Then, the conditional probability of a pixel with RGB values to be skin or non-skin is: $P(rgb|skin) = \frac{Hist_{skin}[r,g,b]}{Total_{skin}}$, and $P(rgb|\sim skin) = \frac{Hist_{non-skin}[r,g,b]}{Total_{non-skin}}$. A new unseen pixel is labeled as skin if it satisfies a given threshold, $\frac{P(rgb|skin)}{P(rgb|\sim skin)} \geq \theta$, where θ is obtained empirically.

The idea behind SPMs seems reasonable given a large amount of training data. Nevertheless, due to the sparse distribution of skin points in RGB space (and memory requirements), one usually reduces the cube's size. This step also helps to "generalize" and compact the histogram. Such generalization in the histogram model is guided by the number of bins. Thus, a SPM also introduces another parameter to fit, i.e., the size and number of bins to consider.

SPM was trained with two-thirds from our dataset and tested on the remaining points. After following the recommendations given in (Brand & Mason, 2000) we got 95.8% for recall, 77.25% for precision, and 91% for success rate[4]. These results are also shown in Table 3 for comparison against RCA.

5. Experimental Results

With the three initial attributes (normalized RGB) and the constant 1/3 we generate, on the first cycle, 39 new attributes with our constructive operators.

Our stopping criterion was set to find a single rule (one path in figure 1) of no more than three simple terms with better success rate than SPM. We restricted the length of the rule to three terms, as we were interested in finding a simple and easy to evaluate rule. This also restricts the number of possible new attributes for the next cycle. If the best found rule is not competitive with SPM the algorithm proceeds to another cycle with the attributes used in the construction of the rules.

The rules were also restricted to cover at least 100 instances from the training set.

The alternative representation spaces produced by RCA are shown in table 3 along with recall, precision, and success rate.

To our surprise, on the first cycle the first representation space of Table 3 had already better precision performance than SPM, is not too far in terms of recall, and has better success rate, even with a much smaller training set. Also, the performance measures of the fifth representation space, shown in table 3, are very similar to SPM. The rule for the first color model is:

[4]Success rate = $\frac{TP+TN}{TP+FP+TN+FN}$, where FN = false negatives.

Figure 2. Performance of the first rule with several images.

$$\begin{array}{lcl} \text{If} & \frac{r}{g} & > 1.185 \text{ and} \\ & \frac{r*b}{(r+g+b)^2} & > 0.107 \text{ and} \\ & \frac{r*g}{(r+g+b)^2} & > 0.112 \\ \text{Then} & \text{Class} = \text{skin} & \end{array}$$

The attributes used by this rule are easy to understand and to evaluate. From experiments (Derrington et al., 1984; Cole et al., 1993) we know that our visual system decomposes colors in three visual pathways (1) luminance, which can be seen as normalizing on $(r+g+b)^2$, (2) a red-green channel, which is present in the first and third attributes, and (3) a blue-yellow, which is not present in our findings. Thus, this color space is suitable for color consistency checking or machine-oriented, rather than human-like. Moreover, the ratio $\frac{r}{g}$ has been widely recognized as an interesting attribute for skin classification (Brand & Mason, 2000; Okada et al., 1989), although it introduces many false positives. Figure 2 shows the performance on this rule on several images.

We then decided to compare our results against C4.5 (Quinlan, 1993) as our decision learner to empirically evaluate the performance of RCA. C4.5 was used to produce rules, instead of RCA in the overall algorithm. We used all the default parameter values for C4.5, except that the rules were forced to cover at least 100 instances of the training set (as in RCA).

C4.5 with the same attributes (39) produced two rules for skin, one with six terms and another one with five terms. This is, in effect, an 11 attributes rule which has just a slightly better performance values than our single rule produced by RCA (93.1% of recall, 94.2%

Table 3. Results of the first level of combinations. These eight models have nine different components, which were selected from a pool of 39. Results from SPM are also shown for comparison.

color model (three components)			recall (%)	precision (%)	success rate (%)
$\frac{r}{g}$	$\frac{rb}{(r+g+b)^2}$	$\frac{rg}{(r+g+b)^2}$	93.7	**91.7**	**92.6**
$\frac{r}{g}$	$\frac{rb}{(r+g+b)^2}$	$\frac{g}{b}$	94.2	88.6	91
$\frac{r}{g}$	$\frac{g}{b}$	$\frac{rb}{(r+g+b)^2}$	94.3	88.5	91
$\frac{b}{g}$	$\frac{r}{g}$	$\frac{rb}{(r+g+b)^2}$	95.1	87.5	91
$\frac{r}{g}$	$\frac{g}{b}$	$\frac{-2r+g+b}{3(r+g+b)}$	95.1	86	90
$\frac{b}{g}$	$\frac{r+g+b}{3r}$	$\frac{r-g}{r+g+b}$	96	85	89.2
$\frac{b}{g}$	$\frac{r}{g}$	$\frac{-2r+g+b}{3(r+g+b)}$	96	84.3	89.1
$\frac{b}{g}$	$\frac{r+g+b}{3r}$	$\frac{r+g-2b}{3(r+g+b)}$	**97.5**	67.1	75
SPM on raw RGB			95.8	77.3	91

of precision and 93.7% of success rate).

$$
\begin{array}{lll}
\text{If} & \frac{g}{r} & \leq 0.839 \text{ and} \\
& \frac{g-b}{r+g+b} & \leq 0.054 \text{ and} \\
& \frac{gb}{(r+g+b)^2} & > 0.067 \text{ and} \\
& \frac{gb}{(r+g+b)^2} & \leq 0.098 \text{ and} \\
& \frac{b}{g} & \leq 1.048 \text{ and} \\
& \frac{g}{3(r+g+b)} & \leq 0.108 \\
\text{Then} & \text{Class} = \text{Skin} &
\end{array}
$$

$$
\begin{array}{lll}
\text{If} & \frac{r+g}{r+g+b} & > 0.685 \text{ and} \\
& \frac{g-r}{r+g+b} & \leq -0.049 \text{ and} \\
& \frac{gb}{(r+g+b)^2} & > 0.067 \text{ and} \\
& \frac{b}{g} & \leq 1.249 \text{ and} \\
& \frac{g}{r+g+b} & \leq 0.324 \\
\text{Then} & \text{Class} = \text{skin} &
\end{array}
$$

It is interested to point out that the attributes selected by C4.5 and RCA are disjoint. Also the attributes of C4.5 are less intuitive. For instance, $\frac{b}{g}$ has been shown to be a bad classifier for pixel based skin detection, as Brand et al. have noticed (Brand & Mason, 2000).

The response time of the models produced by RCA and C4.5 were also evaluated (see Table 4). As can be seen, RCA is on average 50% times faster than C4.5. Reasonably accurate and fast models, as those produced by RCA, are important for any tracking systems, which our main motivation for this research.

Table 4. Response times (in seconds) for 22 million pixels using the models produced by RCA, C4.5, and SPM

	Response Time (secs.)
SPM	4.6
RCA	8.1
C4.5	12.1

In order to increase the performance values of RCA, we allowed it to grow an additional term, producing 16 rules in total. Our best model has the following performance results: 94.1% of recall, 92.7% of precision, and 93.4% of success rate.

$$
\begin{array}{lll}
\text{If} & \frac{b}{g} & < 1.249 \text{ and} \\
& \frac{r+g+b}{3r} & > 0.696 \text{ and} \\
& \frac{1}{3.0} - \frac{b}{r+g+b} & > 0.014 \text{ and} \\
& \frac{g}{3(r+g+b)} & < 0.108 \\
\text{Then} & \text{Class} = \text{skin} &
\end{array}
$$

Although it is still slightly worst than C4.5, it is much easier to evaluate. Allowing RCA to grow a rule with five terms, increases the performance to 94% of recall, 93.6% of precision, and 93.8% of success rate, which is slightly better than C4.5. With six terms, RCA produces several rules of equivalence performance to C4.5 (e.g., one rule has 92.4% of recall, 94.9% of precision, and 93.7% of success rate).

We then decided to run our algorithm for another cycle in search for better rules. We took all the attributes used in the rules of three terms, without repetitions (9 in total) and feed them to our constructive operators. This produced 234 new attributes. The 243 attributes (234+9) were given to RCA which produced the following three terms rule with 93% of recall, 93.5% of precision, and 93.3% for success rate. These results are slightly better than those produced on the first cycle by RCA.

$$
\begin{array}{lll}
\text{If} & \frac{3br^2}{(r+g+b)^3} & > 0.1276 \text{ and} \\
 & \frac{r+g+b}{3r} + \frac{r-g}{r+g+b} & \leq 0.9498 \text{ and} \\
 & \frac{rb+g^2}{gb} & \leq 2.7775 \\
\text{Then} & \text{Class} = \text{skin} &
\end{array}
$$

C4.5 was also run on these 243 attributes, producing three rules (12 terms in total) with 93.5% for recall, 92% for precision, and 92.7% for success rate, which are slightly worst than RCA on the second cycle and than C4.5 on the first cycle. A summary of all these results is given in table 5. It should be noted that RCA has all the best performance measures when compared against C4.5.

$$
\begin{array}{lll}
\text{If} & \frac{g}{b} - \frac{r}{g} & \leq -0.1367 \text{ and} \\
 & \frac{r+g+b}{3*r} - \frac{r*b}{(r+g+b)^2} & > 0.583 \text{ and} \\
 & \frac{r+g+b}{3*r} + \frac{r-g}{r+g+b} & \leq 0.9498 \\
\text{Then} & \text{Class} = \text{skin} &
\end{array}
$$

$$
\begin{array}{lll}
\text{If} & \frac{g}{b} - \frac{r}{g} & \leq -0.0905 \text{ and} \\
 & \frac{g(r+g+b)}{b(r-g)} & > 3.4857 \text{ and} \\
 & \frac{(r+g+b)^3}{3gr^2} & \leq 7.397 \text{ and} \\
 & \frac{r+g+b}{9r} - \frac{1}{3} & > -0.0976 \\
\text{Then} & \text{Class} = \text{skin} &
\end{array}
$$

$$
\begin{array}{lll}
\text{If} & \frac{r+g^2-2b}{3r(r+g+b)} & > 0.014 \text{ and} \\
 & \frac{br^2-rgb}{(r+g+b)^3} & > 0.0075 \text{ and} \\
 & \frac{g(r+g+b)}{b(r-g)} & > 3.4857 \text{ and} \\
 & \frac{r+g+b}{9r} - \frac{1}{3} & > -0.0976 \text{ and} \\
 & \frac{g+b-2r}{r+g-2b} & \leq -1.1022 \\
\text{Then} & \text{Class} = \text{skin} &
\end{array}
$$

Table 5. Summary of results for SPM, RCA, and C4.5.

Algorithm	Recall (%)	Precision (%)	Success Rate (%)
SPM	**95.8**	77.3	91
RCA: 3 terms 1st. cycle	93.7	91.7	92.6
RCA: 4 terms 1st. cycle	94.1	92.7	93.4
RCA: 5 terms 1st. cycle	94	93.6	**93.8**
RCA: 6 terms 1st. cycle	92.4	**94.9**	93.7
RCA: 3 terms 2nd. cycle	93	93.5	93.3
C4.5 11 terms 1st. cycle	93.1	94.2	93.7
C4.5 12 terms 2nd. cycle	93.5	92	92.7

On this second cycle, both RCA and C4.5 produced rules with less intuitive terms and C4.5 decreased its performance. Adding new attributes does not necessarily mean that a learning algorithm will improve its performance. New attributes can bias the learner, at early stages in the construction of rules, towards particular attributes, eliminating better rules in the long run that could have been constructed if such additional attributes were not included. Perhaps better attributes (and bias) could be obtained with a different set of constructive operators. RCA, on the other hand, appears to be less susceptible to the shift of bias that can occur with the addition of new attributes as it searches for rules in parallel. This, however, deserves further tests and is left for future work.

By increasing the number of terms considered on each rule, RCA is able to produce, at least on this domain, much simpler rules with equivalent performance to C4.5. It is clear than further analysis is required in other domains.

Our results for skin detection, either with RCA or C4.5, are to our knowledge, very competitive to any other existing algorithm reported in the literature. Furthermore, the results from RCA are explicit and are simpler to implement than SPM (or C4.5).

6. Conclusions and Future Work

In this paper, we have shown a constructive induction algorithm that uses simple arithmetic operations to change its representation space. The algorithm uses RCA, a restricted covering algorithm, as its selective learner. RCA searches for single rules in parallel, with a predetermined length and using two intermixed criteria for selecting attributes. The constructive induction algorithm obtained better performance results in precision and success rate than SPM, considered as the best skin selection method, and comparable to C4.5, but with considerably simpler models. Fast and accurate models for skin detection are important for human tracking applications.

There are several future research directions that are worth exploring. In particular, the effect of constructing rules with more than one selection criterion needs further analysis. We are assessing the performance on RCA, on several databases, with a single criterion, thus constructing a single rule, and with N (possibly different) criteria, thus constructing N^L rules in parallel (where L is the number of terms in the LHS of the rules). A combined strategy, could be part as well of a general covering algorithm, aiming at producing simple rules. We are also doing a more in-depth analysis of the models found with our strategy from the computer vision point of view.

References

Brand, J., & Mason, J. S. (2000). A comparative assessment of three approaches to pixel-level human skin-detection. *ICPR, Vol. I* (pp. 1056–1059).

Brand, J., Mason, J. S., Roach, M., & Pawlewski, M. (2001). Enhancing face detection in colour images using a skin probability map. *Int. Conf. on Intelligent Multimedia, Video and Speech Processing* (pp. 344–347).

Cole, G. R., Hine, T., & McIlhagga, W. (1993). Detection mechanisms in the l-, m- and s-cone contrast space. *J. Opt. Soc. Am. A* (pp. 38–51).

Derrington, A. M., Krauskopf, J., & Lennie, P. (1984). Chromatic mechanisms in the lateral geniculate nucleus of the macaque. *J. Physiology* (pp. 257–241).

Fleck, M., Forsyth, D. A., & Bregler, C. (1996). Finding nacked people. *ECCV, Vol. II* (pp. 592–602).

Hjelmas, E., & Low, B. K. (2001). Face detection: A survey. *CV&IU, 83*, 236–274.

Jones, M. J., & Regh, J. (1999). Statistical color models with applications to skin detection. *CVPR, Vol. I* (pp. 274–280).

Kumar, V. P., & Poggio, T. (2000). 'learning-based approach to real time tracking and analysis of faces. *Automatic Face and Gesture Recognition* (pp. 96–101).

Langley, P., Bradshaw, G. L., & Simon, H. A. (1983). Rediscovering chemistry with the bacon system. In R. S. Michalski, J. Carbonell and T. M. Mitchell (Eds.), *Machine learning*, 307–329. California: Morgan Kaufmann.

Murphy, P. M., & Pazzani, M. J. (1991). Id2-of-3: Constructive induction of m-of-n concepts for discriminators in decision trees. *Proc. of the Eighth International Workshop on Machine Learning* (pp. 183–187). San Francisco, CA: Morgan Kaufmann.

Okada, K., Oohira, S., & Nakamura, H. (1989). A method of extracting lip shape. *IEICE Trans. on Fundamentals* (pp. 1582–1583).

Pagallo, G. (1990). *Adaptive decision tree algorithm for learning from examples.* Doctoral dissertation, University of California at Santa Cruz, Santa Cruz, CA.

Quinlan, J. (1993). *C4.5: Programs for machine learning.* Morgan Kaufmann.

Ragavan, H., & Rendell, L. (1993). Lookahead feature construction for learning hard concepts. *Proc. of the Tenth International Conference on Machine Learning* (pp. 252–259). San Francisco, CA: Morgan Kaufmann.

Saber, E., & Telkap, A. M. (1998). Frontal-view face detection and facial feature extraction using color, shape and symmetry based cost functions. *Pattern Recognition Letters, 19*, 669–680.

Sigal, L., Sclaroff, S., & Athitsos, V. (2000). Estimation and prediction of evolving color distributions for skin segmentation under varying illumination. *CVPR, Vol. II* (pp. 152–159).

Sobottka, K., & Pitas, I. (1996). Segmentation and tracking of faces in color images. *Automatic face and gesture recognition* (pp. 236–241).

Soriano, M., Martinkauppi, B., Huovinen, S., & Laaksonen, M. (2000). Skin detection in video under changing illumination conditions. *ICPR, Vol. I* (pp. 839–842).

Spors, S., & Rabenstein, R. (2001). A real-time face tracker for color video. *Acoustics, Speech, and Signal Processing, 3*, 1493–1496.

Terrillon, J., Shirazit, M., Fukamachi, H., & Akamatsu, S. (2000). Comparative performance of different skin chrominance models and chrominance spaces for the automatic detection of human faces in color images. *Automatic Face and Gesture Recognition* (pp. 54–61).

Zarit, B., Super, B. J., & Quek, F. K. H. (1999). Comparison of five color models in skin pixel classification. *Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems* (pp. 58–63).

Zheng, Z. (1995). Constructing nominal x-of-n attributes. *Proc. of the Fourteenth International Joint Conference on Artificial Intelligence* (pp. 1064–1070). San Mateo, CA: Morgan Kaufmann.

Zheng, Z. (1998). A comparison of constructing different types of new feature for decision tree learning. In *Feature extraction, construction and selection: A data mining perspective*. Kluwer Academic.