

# 신경회로망 벡터 양자화를 이용한 움직임 탐색 영역의 예측

류대현\*, 남기곤\*\*, 윤태훈, 김재창\*\*

\* : 한국전자통신연구소

\*\* : 부산대학교 전자공학과

**ABSTRACT:** This paper describes a method for estimating motion vectors in a video sequence. In this method, we find motion vectors using the full search method from the training images and then, train the codebook of the neural networks vector quantizer using these motion vectors. A motion vector can be estimated using the codebook as a motion prediction region. The codewords in the codebook represent the motion vectors for the input image sequences. Since the codebook is used as the search region for estimating the motion vectors, search points and computation can be reduced compared with the full search block matching algorithm. Additionally, the information required to transmit the motion vectors can be reduced.

## 1. 서론

화상 회의, 화상 전화, 디지털 HDTV 등에서 동영상 압축 기술은 매우 중요한 부분을 차지한다. 동영상에서는 공간 상관성 뿐만 아니라 시간 상관성도 매우 크기 때문에, 이를 이용하여 시간축의 중복성을 제거하는 방법인 움직임 보상 부호화(motion compensated coding: MCC)를 이용하면 높은 데이터 압축율을 얻을 수 있다. 이러한 동영상 부호화 방식에 있어서 이동 정보량의 추정은 매우 중요한 문제이며 많은 연구가 이루어져 왔다[1-4].

움직임 보상 부호화는 시간상의 중복성을 제거하기 위해 먼저 정확한 움직임 추정(motion estimation)에 의해 움직임을 보상하여 시간상의 중복성을 제거하는 부분과 그 예측 오차를 부호화하는 부분으로 이루어져 있다. 움직임 추정 방법에는 크게 화소 순환 알고리즘(pel recursive algorithm: PRA)과 블럭 정합 알고리즘(block matching algorithm: BMA)이 있다[3]. 이 중 블럭별로 움직임 벡터를 찾아 블럭내의 모든 화소에 같은 움직임을 적용하는 블럭 정합 알고리즘은 실시간 하드웨어 구현이 용이하기 때문에 HDTV(high definition television) 및 MPEG(moving picture experts group)의 비디오 압축 알고리즘 등에 널리 사용되고 있다[1,2].

블럭 정합 알고리즘을 근간으로 하는 움직임 보상 예측 부호화 방법에서는 움직임 벡터와 예측 오차가 함께 부호화 된다. 이때 효율적인 부호화 방법을 사용함으로써 움직임 벡터와 예측 오차가 가능한 적은 정보량을 갖게 해주는 것이 매우 중요하다. 움직임 벡터에 대한 정보량은 사용하는 부호화 기법 및 전송률에 따라 다르다. 예를 들어 384Kbits/s의 전송율을 갖는 코덱(codec)에서 전체 정보량의 약 40%가량을 차지하는 경우도 있다. 이렇게 움직임 벡터가 많은 정보량을 갖고 있는 경우에는 움직임 벡터의 효율적인 부호화는 매우 중요한 문제가 된다[3,4].

본 논문에서는 움직임 벡터를 추정하기 위해 신경 회로망을 이용한 벡터 양자화에 의해 탐색 영역을 예측하고 예측 점에 대해서만 왜곡을 계산함으로써 움직임 벡터를 구하는 방법을 제안한다. 시험 영상에서 전역 탐색(full search) 법을 이용하여 움직임 벡터를 구하고 이를 입력으로 FSCL 신경 회로망 벡터 양자화기의 코드 북을 설계하고 이것을 움직임 예측 영역으로 이용하여 블럭 정합 알고리즘을 수행하여 움직임 벡터를 구한다. 이렇게 함으로써 결과적으로 영상에서 움직임 벡터의 공간적 상관성을 이용하게 되어 탐색점의 수 즉, 계산량을 줄일 뿐만 아니라 움직임 벡터 전송을 위한 소요 비트 수를 줄임으로써 결과적으로 압축율을 높인다. 모의 실험 결과 제안된 방법은 PNSR 측면에서 비교적 성능이 우수한 고속 정합 알고리즘으로 알려진 3단계 탐색법(three steps search:TSS)에 의한 것보다 향상됨을 보인다[5].

움직임 예측 영역을 나타내는 코드북을 학습시키는 데에는 많은 계산량이 요구된다. 그러나 이러한 계산은 인코딩되지 않은 비디오에 대해 오프 라인, 또는 비교적 저속의 디지털 신호 처리기 등에 의해 병렬로 이루어 질 수 있으며, 망의 상태에 따라 코드 북의 크기, 즉 탐색 점의 수를 적응적으로 변화 시킴으로써 비트율을 제어할 수 있다.

이러한 방식은 폭주 제어 방식을 사용하는 비 동기 전송 모드에서 요구형 비디오와 같은 응용

분야에 적용될 수 있다. 이러한 응용 분야에서는 코딩되지 않은 영상 열이 송신단에 항상 저장되어 있으며 수신 장치에서 요구가 있을 때 송신 서버는 요구된 영상 열을 인코딩하여 송신한다. 비디오 코더는 망으로부터 망의 상황에 대한 피드백 정보를 수신하여 망에 폭주가 발생하였을 때는 패킷 손실을 줄이기 위해 출력율을 자동으로 줄인다. 이러한 적응 방식은 우선 순위에 따라 패킷을 드롭시키는 방식에 비해 상대적으로 우수한 성능을 보인다[6,7].

본 논문의 구성은 다음과 같다. 2장에서 기존의 블럭 정합 알고리즘을 간단히 소개하고 그 문제점을 설명한다. 3장에서 신경 회로망 벡터 양자화를 이용하여 움직임 영역을 예측하고 이를 이용하여 움직임 벡터를 검출하는 하는 방법을 기술한다. 4장에서는 기존의 알고리즘과 제안한 방법을 모의 실험을 통해 그 결과를 비교 검토하고 5장에서 결론을 맺는다.

## 2. 블럭 매칭 알고리즘을 이용한 움직임 벡터의 추정

블럭 매칭 알고리즘은 영상 열에서 연속하는 프레임 간에서 물체의 이동량을 검출하기 위하여 부영상 사이의 상관 관계를 비교하여 최대치를 이동 보상 위치로 이용하는 방법이다[8]. 그림 1에 블럭 매칭 알고리즘에 의한 움직임 검출 방식이 나타나 있다.

블럭 매칭 알고리즘의 처리 과정은 다음과 같다. 우선 화상을 고정된 크기의 부영상으로 나눈다. 이때 이전 화상내의 부영상과의 상관계수 값이 최대가 되는 위치(오차가 제일 작게 일어나는 위치)를 구하기 위해

$$D(i, j) = \frac{1}{NM} \sum_{m=1}^M \sum_{n=1}^N G(U(m, n) - U_r(m+i, n+j)) \quad (1)$$

$$-p \leq i, j \leq p$$

와 같은 함수  $D(\cdot)$ 을 정의한다. 여기서,  $G(\cdot)$ 은 오차 전력(error power)을 구하는 비선형 함수,  $U$ 는 현 영상내에서  $M \times N$  크기의 부영상으로 구성된 블럭,  $U_r$ 은 이전 화상 내에서  $(M+2p)(N+2p)$  크기의 탐색 영역,  $p$ 는 최대 이동 가능 거리이다. 이때 이동치는  $D(i, j)$ 를 최소로 하는  $(i, j)$ 로 주어진다.  $D(\cdot)$  값은 최적 위치에서 멀어질수록 증가한다. 즉 정확한 이동치를 갖는 위치에서 멀어지면 멀어질수록 오차가 커지므로 이동값을 구하는 것은  $D(\cdot)$  값을 적게하는 위치로 찾아 가는 것이다. 이와 같은 가정하에서 계산량을 줄일 수 있을 뿐아니라

탐색 속도를 향상시키기 위한 블럭 정합 방법에 대한 많은 연구가 이루어 졌으며, DMD(direction of minimum distortion), TSS 그리고 OTS(one at a time search) 등의 방법이 제안되었다[9,10].

검출된 움직임 벡터를 이동 보상형 예측 또는 이동 보상형 내삽 등에 이용하기 위해서는 운동체의 실제 이동 정보의 추출, SNR 등의 예측 성능, 전송될 이동 벡터의 데이터 량 등이 고려되어야 한다[11,12]. 위와 같은 측면을 고려할 때 블럭 매칭 알고리즘에 의한 움직임 벡터 검출 방식은 널리 사용되고 있음에도 불구하고 많은 단점을 갖고 있다. 즉 블럭 대 블럭으로 운동벡터를 검출하므로 그 블럭내에서 움직임이 균일하다는 가정이 필요하다. 이러한 가정은 상대적으로 작은 블럭( $8 \times 8$  또는  $16 \times 16$ )에서만 만족한다. 그러나 블럭 크기가 작아진다는 것은 블럭의 수가 늘어나고 전송해야 할 운동 벡터가 늘어나는 결과가 되므로 전송 부담이 생긴다. 또한 블럭 매칭 알고리즘을 적용하여 구해진 이동 벡터는 운동체의 국지적인 특성만을 이용하며 인접 벡터간들의 연관성을 고려하지 않았기 때문에 운동체의 실제 이동과 일치하지 않는 경우가 발생한다.

이런 문제점을 해결하기 위해 움직임이 일어나는 경계 부분에 대해서 보다 섬세한 움직임 추정을 하기 위하여 블럭 크기를 가변시키는 등 블럭 매칭 알고리즘에 대한 많은 연구가 이루어 졌다. 그러나 블럭 크기를 크게 하여 움직임을 추정 한 후 예측 오차가 크면 블럭을 나누는 방식에서는 중복된 계산으로 인해 효율적이지 못하고, 또한 블럭 크기가 너무 다양한 경우에는 이동 벡터 량의 증가로 인해 부호화시키기가 복잡하게 된다는 등의 단점을 피할 수 없다.

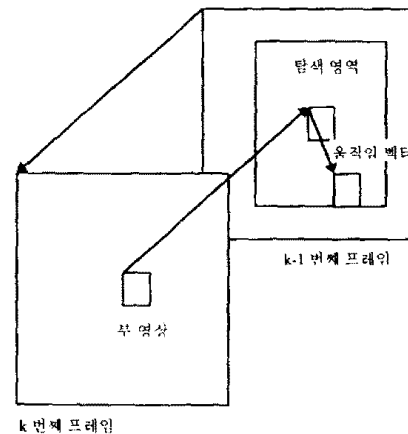


그림 1. 블럭 매칭 알고리즘에 의한 움직임 검출.

Fig. 1. Motion detection by block matching algorithm

### 3. 신경 회로망 벡터 양자화를 이용한 움직임

#### 임 벡터의 추정

일반적으로 블럭 매칭 알고리즘에 의해 찾아진 움직임 벡터들은 높은 시공간 상관성을 가지고 있기 때문에 이를 이용하여 움직임 벡터 검출의 성능을 높힐 수 있다. 본 장에서는 훈련 영상 열의 초기 프레임에서 전역 탐색법을 이용하여 움직임 벡터를 구하고, 이를 입력으로 신경 회로망을 학습하여 코드북을 생성하고 이를 움직임 예측 영역으로 이용하여 움직임 벡터를 추정하는 방법을 제안한다. 이렇게 함으로써 영상에서 움직임 벡터의 공간적 상관성을 이용하여 움직임 벡터 추정시 탐색점의 수를 줄임으로써 계산량을 줄이고 움직임 벡터 전송을 위한 부가 정보의 양을 줄여 데이터 압축율도 높힐 수 있다.

벡터 양자화는 입력 데이터에 내재하는 통계적 특성을 이용하는 양자화 기법으로서 입력 벡터 공간을 여러 개의 영역으로 나누고 각 영역에 대해 대표 벡터를 계산한다. 이들 대표 벡터를 코드어라 하며 그 집합을 코드북이라 하는데 코드북이 생성된 다음, 양자화될 입력 데이터 벡터가 주어지면 그 벡터가 속한 영역을 찾고 그 영역에 대한 대표 벡터의 인덱스에 의해 그 입력 데이터 벡터를 표현한다. 다시 말하면 벡터 양자화란 적절한 거리 척도에 의해 이미 저장된 기준 벡터 중에서 가장 잘 일치하는 벡터로 매핑시켜주는 코딩 방식이라 할 수 있다.

움직임 벡터들이 공간적 상관성을 갖는 일련의 영상 프레임 열의 움직임 벡터의 양자화에 이러한 벡터 양자화를 적용할 경우 훈련에 의해 생성된 코드북은 대표 움직임 성분(representitive motion component)을 나타내며, 코드북 설계시 코드북의 크기를 적당히 조절하면 움직임 벡터 예측 오차에 의한 손실을 최소화하며 움직임 벡터의 데이터량을 줄일 수 있다. 즉 코드북의 크기를 줄이면 탐색 영역이 줄어들므로 계산량이 적어지고 압축비를 높일 수 있으나, 움직임 예측 오차가 증가하므로 잉여신호의 전송시 데이터량이 증가할 뿐 아니라 화질의 저하를 초래할 수도 있다. 반면 코드북의 크기를 크게 하면 탐색 영역이 늘어나므로 계산량이 증가하고 전송해야 할 움직임 벡터의 정보량이 증가한다. 따라서 ATM 망 등에서 망의 상태에 따라 적응적으로 코드북의 크기를 조절함으로써 성능을 최적의 상태로 유지하는 데 이용될 수 있다[7].

일반적으로 벡터 양자화를 위한 코드북을 생성하기 위한 훈련 과정과 인코딩 과정은 매우

복잡하고 계산량도 많다. 또한 LBG 알고리즘과 같이 현재 사용되는 대부분의 알고리즘은 배치 모드 알고리즘이며, 훈련 과정에서는 훈련 데이터를 모두 액세스 해야할 필요가 있다. 또한 입력 데이터의 통계적 특성이 상황에 따라 변할 수 있는 경우에는 입력이 들어올 때 마다 코드북을 갱신하는 적응 벡터 양자화 방식을 사용하는 것이 유용하다. 이러한 적응 벡터 양자화에는 많은 병렬 처리 특성과 다양한 학습 알고리즘들을 적용할 수 있는 신경 회로망을 사용하는 것이 매우 유리하다.

비교사 경쟁학습 알고리즘을 갖는 신경 회로망으로 벡터 양자화를 수행하면 신경 회로망의 병렬 처리 구조로 인해 실시간 처리가 가능하고 신경 회로망의 다양한 학습 방법에 의해 벡터 양자화 알고리즘이 개선될 수 있는 장점이 있다 [13].

본 논문에서 제안한 신경 회로망 벡터 양자화기를 이용한 움직임 벡터 추정 방식의 블럭도는 그림 2와 같다. 먼저 두 프레임의 시험 입력 영상에 대해 전역 탐색법에 의해 움직임 벡터를 구한다. 실시간으로 계산되지 않아도 되므로 보다 정교한 움직임 벡터 추정 방법을 사용한다면 더 좋은 결과를 얻을 수 있을 것이다. 이렇게 구해진 움직임 벡터를 훈련 벡터로 이용하여 신경 회로망을 훈련하여 코드북을 생성한다. 최종 생성된 코드북의 코드어들은 시험 입력 영상 열의 움직임 벡터들을 대표한다. 영상 열에 대해서 이 코드북을 탐색점으로 사용하여 움직임을 추정한다면 전역 탐색법에 비해 탐색점의 수와 계산 시간을 줄일 수 있다.

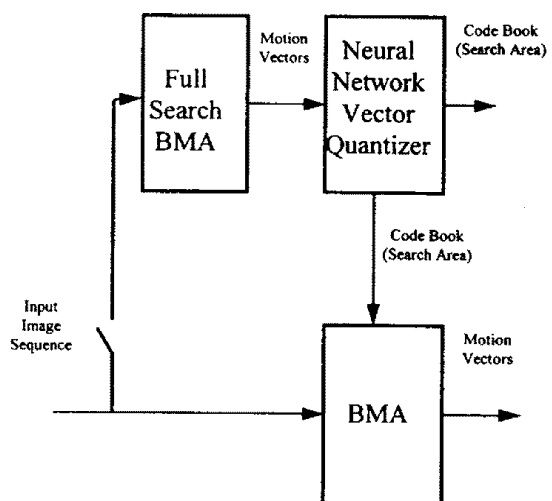


그림 2. 제안된 움직임 벡터 추정 방식 블럭도.

Fig. 2. Block diagram of suggested motion vector estimation method.



그림 3. 전역 탐색에 의해 구해진 시험 영상의 움직임 벡터와 그 분포(블록 크기=8, 탐색 영역=±7).

Fig. 3. Motion vector and it's distribution by full search(block size=8, search area=±7).

또한 움직임 벡터를 전송하는 대신에 인덱스를 전송하므로 움직임 벡터를 전송하기 위한 정보량도 줄어들게 된다. 예를 들어 ±7의 탐색 영역을 갖는 블록 매칭 알고리즘을 사용할 경우 가능한 움직임 벡터의 수는 225가지가 되고 이를 전송하려면 고정 길이로 인코딩하는 경우 움직임 벡터당 8비트가 필요하지만 제안한 방법에 의해 코드 북의 크기를 25로 설계한다면 5비트만 필요하게 되므로 압축 비를 높일 수 있다. 또한 탐색점의 수가 225점에서 25점으로 줄어들므로 계산량의 측면에서도 성능 개선이 이루어진다.

그림 3에 시험 영상 열의 초기 2 프레임을 사용하여 전역 탐색법에 의해 움직임 벡터를 구한 결과가 나타나 있다. 이 그림에서 동일한 영상 열의 움직임 벡터들은 큰 공간적 상관성을 가지고 있음을 알 수 있다. 이들 움직임 벡터들을 훈련 입력 데이터로 하여 앞 절에서 설명된 신경 회로망 벡터 양자화기에서 FSCL(frequency sensitive competitive learning) 학습 알고리즘에 의해 코드 북을 생성한다.

#### 4. 모의 실험 및 고찰

모의 실험에서는 'Flower Garden'(그림 3) 30 프레임을 시험 영상으로 사용하였다. 이 영상은 CCIR601의 1/2의 크기를 가지는 SIF 형태이며 8비트로 양자화되어 있다. 블록 매칭 알고리즘에서의 블록 크기는 8×8로 하였다. MPEG에서는 탐색 영역을 수직과 수평 방향으로 각각 ±15를 권

고하였으나 실험에서 사용한 시험영상이 CCIR601의 1/2의 크기이므로 탐색 영역을 ±7로 하여 실험하였다.

신경 회로망 벡터 양자화기를 이용하여 탐색 영역을 예측하기 위해서는 코드 북을 훈련하기 위한 시험 영상이 필요한데 본 실험에서는 초기 두 프레임으로부터 구해진 움직임 벡터를 사용하였다. 코드북의 크기 즉, 탐색점의 수는 25로 하였다.

표 1에 나타난 비와 같이 고속 블록 매칭 방식 중 성능이 우수한 것으로 알려진 TSS의 경우 탐색 영역의 크기는 27 점이 되며, 전역 탐색법의 경우는 225점이 된다. 따라서 본 논문에서 제안한 방식은 탐색점의 수를 줄이므로 계산량을 감소시킬 뿐만 아니라 움직임 벡터를 전송하는 데 필요한 비트수도 크게 줄어든다.

제안된 방식의 경우는 전역 탐색법에 의한 블록 매칭 방식으로 움직임 벡터를 검출하는 경우와 비교하면 ±2 전역 탐색 블록 매칭 방식에 해당되므로 이 경우의 결과도 나타내었다.

제안된 방법은 전역 탐색법과 비교하여 평균 PSNR 측면에서 거의 차이가 없으나 삼단계 탐색법에 비하여 1.5dB 이상 우수함을 보여 준다. 계산량과 움직임 벡터를 전송하기 위한 부가 정보량의 측면에서 유사한 ±2 전역 탐색 블록 매칭 방식과 비교할 때 5dB 이상 우수함을 보여 준다.

그림 4는 flower garden 30 프레임에 대한 PSNR 그래프이다. flower garden과 같이 움직임 벡터의 상관성이 매우 큰 영상열에 대해서는 제안된 방식이 TSS보다 위c등히 우수하며 거의 전역 탐색법과 근사한 성능을 보여주고 있다.

또한 제안된 방법은 움직임 벡터를 벡터 양자화하는 과정에서 움직임 벡터를 스무딩하는 것과 같은 효과를 얻을 수 있다. 이러한 스무딩 효과는 움직임 벡터의 양자화 과정에서 발생하는 것으로 잡음 등에 의해 움직임 벡터를 잘못 추정하는 것을 방지하고 움직임 벡터의 엔트로피 부호화 효율도 높인다.

표 1. 성능 비교.

Table I. Performance comparison.

method	number of search point	required bits per a motion vector	average PSNR(dB)
Full search(±7)	221	8 bits	22.6791
Full search (±2)	25	5 bits	17.3795
TSS	27	5 bits	21.0388
Proposed	25	5 bits	22.6388

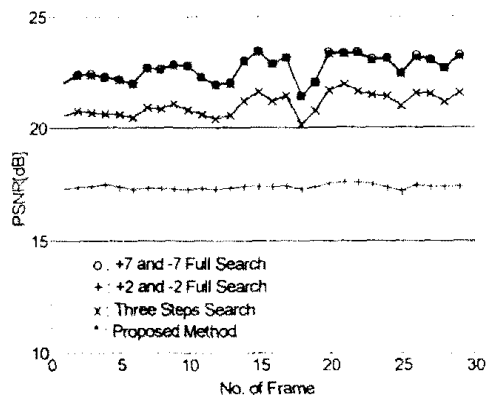


그림 4. Flower garden 영상 열에 대한 성능 비교.

Fig. 4. Performance comparison for flower garden image sequence.

## 5. 결론

본 논문에서는 초기 영상에서 전역 탐색 블록 매칭 알고리즘을 이용하여 움직임 벡터를 구하고 이를 훈련 움직임 벡터로 FSCL 신경 회로망에 적용하여 코드북을 설계하였으며 이 코드북을 움직임 예측 영역으로 이용하였다.

이러한 방법은 영상에서 움직임 벡터의 공간적 상관성을 이용한 것으로써 블록 매칭 알고리즘에서 탐색 영역 즉 계산량을 줄이고 움직임 벡터 전송을 위한 소요 비트를 줄임으로써 압축율을 높인다.

모의 실험 결과 제안된 방법은 PSNR 측면에서 삼단계 탐색법에 의한 것보다 평균 1.5dB 이상의 향상되는 결과를 보였다.

제안된 방법은 움직임 벡터를 벡터 양자화하는 과정에서 움직임 벡터를 스무딩하는 것과 같은 효과를 얻기 때문에 잡음에도 강한 특성을 보였다.

## 참고문헌

[1] ISO-IEC/JTC1/SC29/WG11 MPEG 92/086, "Preliminary working draft of test model 0," Mar. 1992.

[2] ISO-IEC/JTC1/SC2/WG8/N MPEG 90, "MPEG video simulation model three(SM3)," 1990.

[3] K. Inuma, T. Koga, K. Niwa, and Y. Iijima, "A motion-compensated interframe codec", *In Proc. Image Coding, SPIE*, vol. 594, pp. 194-201, 1985.

[4] 김기현, 김진태, 장태규, 최종수, "효율적인 부호화와 예측오차를 고려한 움직임 벡터스무딩 기법에 관한 연구," *제5회 신호처리 합동 학술대회 논문집*, pp.719-723, 1992.

[5] D. H. Ryu, C. R. Kim, T. W. Choi, J. C. Kim, "New search region prediction method for motion estimation," *3rd International Workshop in Signal and Image Processing, Manchester*, Nov. 1996.

[6] Y. Y. Lee and J. W. Woods, "Motion vector quantization for video coding," *IEEE Trans. Image Processing*, vol 4, no. 3, Mar. 1995.

[7] H. Kanakia, P. Mishra, and A. Reibman, "An adaptive congestion control scheme for real-time packet video transport," *In ACM sigcomm'93 Conf. Proc.: Commun. Architectures, Protocols, Applicat.*, San Francisco, pp. 20-31, Sep. 1993.

[8] F. Giorda and A. Racciu, "Bandwidth reduction of video signal via shift vector transmission," *IEEE Trans. Commun.*, vol. COM 23, no.3, pp. 1002-1003, Mar. 1977.

[9] T. Koga et al., "Motion compensated interframe coding for video conferencing," *Nat. Telecom. Conf.*, Nov. 1981.

[10] 이인홍, 박래홍, "이동벡터 추정을 위한 고속 block matching algorithm에 관한 연구," *대한 전자공학회지*, 제25권, 제2호, pp.93-101, 1988.

[11] S. Ullman, "Analysis of visual motion by biological and computer systems," *IEEE Computer*, vol. 14, no. 8, pp.57-69, 1981.

[12] M. Bierling, "Displacement estimation by hierarchical block matching," *Proc. of SPIE : Visual Communications and Image Processing*, vol. 1001, pp.942-951, 1988.

[13] S. C. Ahalt, A. K. Krishnamurthy, P. Chen and D. E. Melton, "Competitive learning algorithms for vector quantization," *Neural Networks*, vol. 3, no. 3, pp. 277-290, 1990.