Abs ac—Rival penalized competitive learning (RPCL) has been shown to be a useful tool for clustering on a set of sample data in which the number of clusters is unknown. However, the RPCL algorithm was proposed heuristically and is still in lack of a mathematical theory to describe its convergence behavior. In order to solve the convergence problem, we investigate it via a cost-function approach. By theoretical analysis, we prove that a general form of RPCL, called distance-sensitive RPCL (DSRPCL), is associated with the minimization of a cost function on the weight vectors of a competitive learning network. As a DSRPCL process decreases the cost to a local minimum, a number of weight vectors eventually fall into a hypersphere surrounding the sample data, while the other weight vectors diverge to infinity. Moreover, it is shown by the theoretical analysis and simulation experiments that if the cost reduces into the global minimum, a correct number of weight vectors is automatically selected and located around the centers of the actual clusters, respectively. Finally, we apply the DSRPCL algorithms to unsupervised color image segmentation and classification of the wine data.

Inde Te ms—Clustering analysis, competitive learning (CL), convergence, cost function, gradient descent.

A ... k ...

... k ... k ... k ... k ...

$\mathcal{G}$

$W_1^{(0)}, \ldots, W_n^{(0)}$ $\mathcal{G}$

Separation Nature $\mathcal{G}$ $\mathcal{G}$ $\mathcal{G}$ $\mathcal{G}$ $\mathcal{G}$

Correct Division $\mathcal{G}$ k

Correct Location k k

$\mathcal{G}$

A. D C F

$\mathcal{S} = \{X^\mu\}_{\mu=1}^N$ $X^\mu = [x_1^\mu, x_2^\mu, \ldots, x_d^\mu]^{\mathrm{T}}$ k

$$E_{\mathrm{MSE}}(\mathbf{W}) = \frac{1}{2}\sum_{ij\mu} M_i^\mu \left(x_j^\mu - w_{ij}\right)^2 = \frac{1}{2}\sum_{i\mu} M_i^\mu \|X^\mu - W_i\|^2 = \frac{1}{2}\sum_{\mu} \left\|X^\mu - W_{c(\mu)}\right\|^2$$

$\mathbf{W} = [W_1, W_2, \ldots, W_k]$ $n = k$ $W_i = [w_{i1}, w_{i2}, \ldots, w_{id}]^{\mathrm{T}}$ i

$$M_i^\mu = \begin{cases} 1, & i = c(\mu) \\ & \left\|X^\mu - W_{c(\mu)}\right\| = \min_j \|X^\mu - W_j\| \\ 0, & \end{cases}$$

$c(\mu)$ μ X^μ $c(\mu)$ M_i^μ X^μ X^μ i X^μ i $E_{\mathrm{MSE}}(\mathbf{W})$ k

$$E(\mathbf{W}) = E_1(\mathbf{W}) + E_2(\mathbf{W})$$

$$E_1(\mathbf{W}) = E_{\mathrm{MSE}}(\mathbf{W}) = \frac{1}{2}\sum_{\mu} \left\|X^\mu - W_{c(\mu)}\right\|^2$$

$$E_2(\mathbf{W}) = \frac{2}{P}\sum_{\mu, i \neq c(\mu)} \|X^\mu - W_i\|^{-P}$$

$\mathbf{W} = \mathrm{vec}[W_1, W_2, \ldots, W_n]$ P

$c(\mu) = 1$

$c(\mu) = 0$ $E(\mathbf{W})$ $\mathbf{W}$ $E(\mathbf{W})$ n $n = 1$ n k

B. D C F

X^μ $\mathcal{B}$ $\mathbf{W}$ $\mathcal{B}$ $\mathcal{B}$

$\mathbf{W}$ $\mathbf{W}$ $\mathcal{B}$

$\mathbf{W}$ X^μ M_i^μ $\mathbf{W}$ $E(\mathbf{W})$ $\mathbf{W}$ w_{ij}

$$\begin{aligned}\frac{\partial E(\mathbf{W})}{\partial w_{ij}} &= \frac{\partial E_1(\mathbf{W})}{\partial w_{ij}} + \frac{\partial E_2(\mathbf{W})}{\partial w_{ij}} \\ &= -\sum_{\mu} \delta_{i,c(\mu)} \left(x_j^\mu - w_{ij}\right) + \sum_{\mu,i} \left(1 - \delta_{i,c(\mu)}\right) \\ &\quad \times \|X^\mu - W_i\|^{-P-2} \left(x_j^\mu - w_{ij}\right)\end{aligned}$$

$\delta_{i,j}$

$\mathbf{W}$ X^μ $\mathbf{W}$ R^{nd} M_i^μ $\mathbf{W}'$ W_i' W_j' $i < j$ $X^{\mu'}$

$$\left\|W_i' - X^{\mu'}\right\| = \left\|W_j' - X^{\mu'}\right\| = \min_l \left\|W_l' - X^{\mu'}\right\| > 0.$$

$\mathbf{W}'$ $\|W_i - X^{\mu'}\| = \|W_j - X^{\mu'}\|$ $\mathcal{A}_{l_i}$ $\mathcal{A}_{l_j}$ $\|W_i - X^{\mu'}\| = \min_l \|W_l - X^{\mu'}\| \leq \|W_j - X^{\mu'}\|$ $\mathcal{A}_{l_i}$ $\|W_j - X^{\mu'}\| = \min_l \|W_l - X^{\mu'}\| < \|W_i - X^{\mu'}\|$ $\mathcal{A}_{l_j}$ $i < j$ $\mathcal{A}_{l_i}$ $\mathcal{A}_{l_i}$ $\mathbf{W}'$ $M_i^{\mu'} = 1$ $M_j^{\mu'} = 0$ $\mathcal{A}_{l_j}$ $\mathbf{W}'$ $M_i^{\mu'} = 0$ $M_j^{\mu'} = 1$ $\|W_i' - X^{\mu'}\| = \|W_j' - X^{\mu'}\|$

$\mathcal{S}$ n

$$C_i = \mathcal{S} \cap R_i, \qquad i = 1, \ldots, n.$$

X^μ $\overline{C}_i = \mathcal{S} - C_i$ C_i

$E_1(\mathbf{W})$

$E_2(\mathbf{W})$

$E_2(\mathbf{W})$ X^μ $\|X^\mu - W_{r(\mu)}\|^{-P}$ $r(\mu)$

$E_2(\mathbf{W})$

$$\Delta W_i = \begin{cases} \eta(X^\mu - W_i), & i = c(\mu) \\ -\eta \|X^\mu - W_i\|^{-P-2}(X^\mu - W_i), & i = r(\mu) \\ 0, & \end{cases}$$

$$\alpha_c = \eta, \qquad \alpha_r = \eta \left\|X^\mu - W_{r(\mu)}\right\|^{-P-2}.$$

$$\frac{\alpha_c}{\alpha_r} = \left\|X^\mu - W_{r(\mu)}\right\|^{2+P}.$$

α_c/α_r $\|X^\mu - W_{r(\mu)}\|$ P α_c α_r η

η η η

η η

$\mathbf{W}^{(t)} = [W_1^{(t)}, \ldots, W_n^{(t)}]$

$\mathbf{W} = [W_1^{(0)}, \ldots, W_n^{(0)}] \in R^{nd} - \mathcal{B}$

t

1: $W_i^{(t)}$

t

: $W_i^{(t)}$

t

$$\Delta W_i = \eta \sum_{\mu} \|X^{\mu} - W_i\|^{-P-2}(W_i - X^{\mu}).$$

ΔW_i

$W_i^{(t+1)}$,

$E(\mathbf{W}^{(t)})$

$E(\mathbf{W}^{(t)})$

W_j

$W_i^{(t)}$

$W_i^{(t+1)}$

W_j W_j

$W_i^{(t+1)}$ $W_i^{(t+1)}$

$W_i^{(t+1)}$

$W_i^{(t+1)}$

W_i

■

1:

η $\mathcal{G}$

$W_i^{(t)}$ $\mathcal{G}$ $\mathcal{G}$

:

$E(\mathbf{W}^{(t)})$ t

η

$E(\mathbf{W}^{(t)})$

E^*

t $\{W_i^{(t)}\}$

$\{W_i^{(t)}\}$

$W_i^{(t)}$

$E(\mathbf{W}^{(t)})$

E^* $\{W_i^{(t)}\}$

$W_i^{(t)}$

$\{W_i^{(t)}\}$ $\mathcal{G}$

$\{W_i^{(t)}\}$

T $t > T$ $W_i^{(t)}$ $\mathcal{G}$

■

η

$E(\mathbf{W})$

$\hat{\mathbf{W}}^*$ $E(\hat{\mathbf{W}})$

$\hat{\mathbf{W}}$

$\hat{\mathbf{W}}^*$

$\hat{\mathbf{W}}^*$

η

$E(\hat{\mathbf{W}})$

$E(\hat{\mathbf{W}})$

$\|W_i - W_j\| \geq \delta$ $i \neq j$ δ

$E(\mathbf{W})$

$\{\mathbf{W} : E(\mathbf{W}) \leq C\}$

$C > 0$ $E(\mathbf{W})$

η

$E(\mathbf{W}^{(t)})$

B. C D

k

k

$n-k$

k

k

k

$m_1,\dots,m_k$

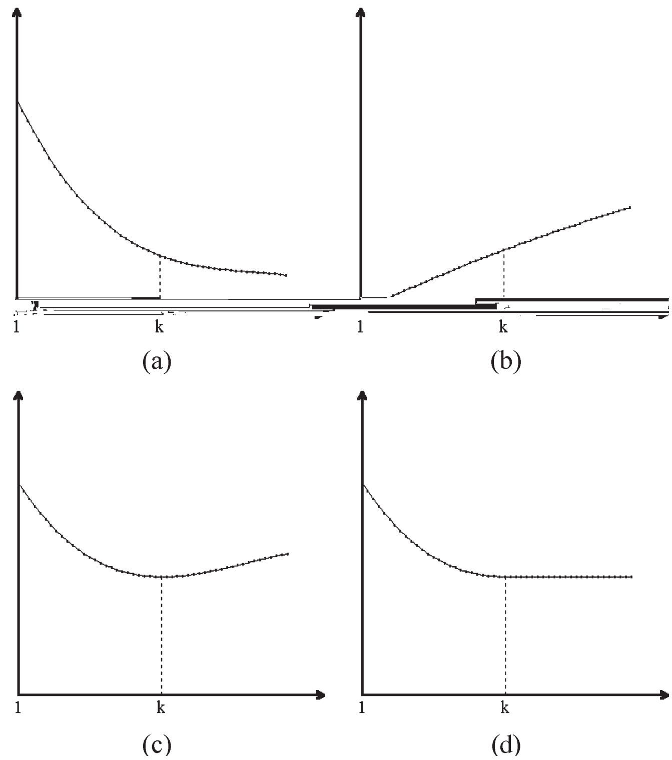
m_i

1) $\mathbf{W}^*$ $E(\mathbf{W})$ $n=k$:

$E(\mathbf{W}) =$
$E_1(\mathbf{W}) + E_2(\mathbf{W})$
$E_1(\mathbf{W})$
$\mathbf{W}^0$
$\mathbf{W}^0$

$m_1,\dots,m_k$
$\mathbf{W}^0 = [m_1,\dots,m_k]$

$\mathbf{W}$ k $C_1,\dots,C_k$
$E_1(\mathbf{W})$ C_i
$E_1(\mathbf{W})$ $E_1(\mathbf{W}^0)$ $\mathbf{W}$
$\mathbf{W}^0$



n k $E_1(\mathbf{W})$ $E_2(\mathbf{W})$ $E(\mathbf{W})$ $E(\mathbf{W})$

$E(\mathbf{W})$

$E(\mathbf{W}^{(t)})$ $E(\mathbf{W})$

$E(\mathbf{W})$ $n = k$

$E(\mathbf{W})$

C. P

$E(\mathbf{W})$. $E(\mathbf{W})$

$n = k$ k

P

P

P

P

P

$[P_0, P_1]$ P

$P \geq 0.01$ $E(\mathbf{W})$ $E_1(\mathbf{W})$ $n = k$

P_0 $P > 1.9$

P_1 P P

A. *E* *C* *A*

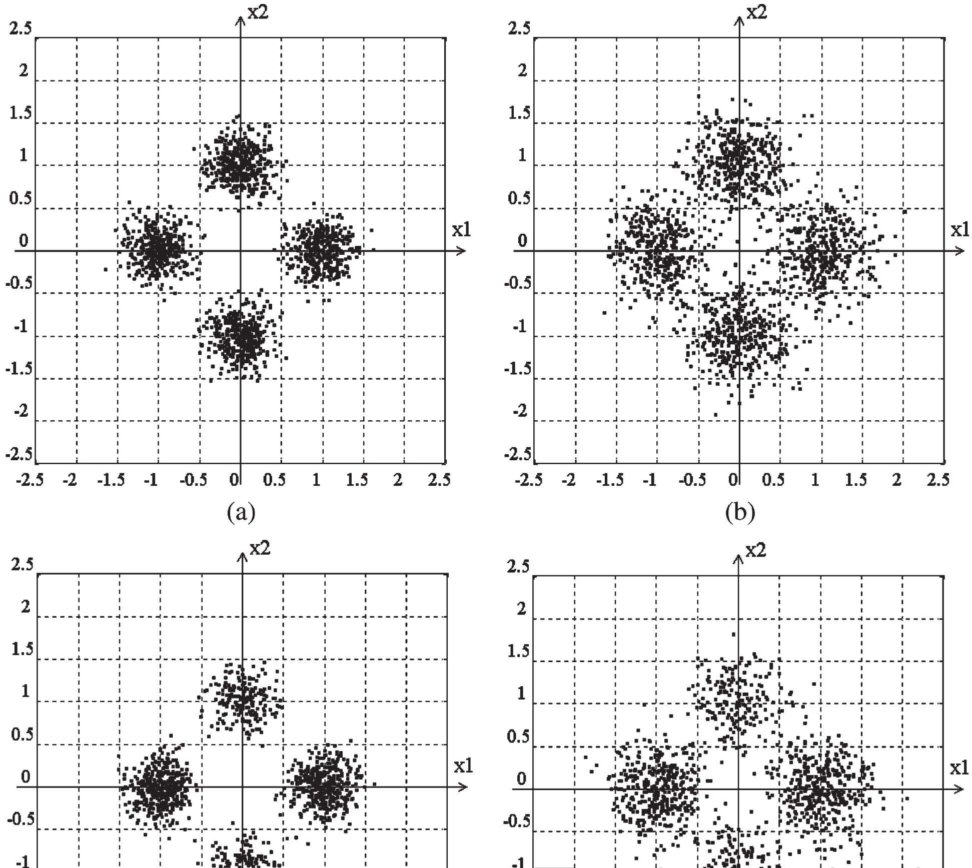
1) *D*:

$d = 2$

$\sigma^2 I$ σ

m_i σ_i α_i N_i

i



$\mathcal{S}_1$ $\mathcal{S}_2$ $\mathcal{S}_3$ $\mathcal{S}_4$ $\mathcal{S}_5$

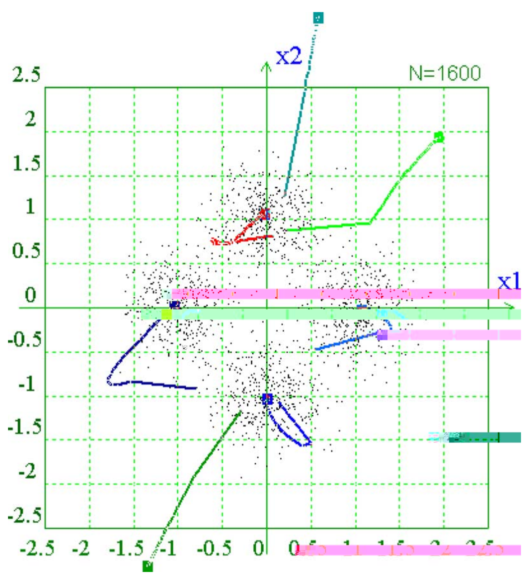
2) D C A :

$E(\mathbf{W})$

$E(\mathbf{W}^{(t)})$

$E(\mathbf{W})$ n

[redacted]



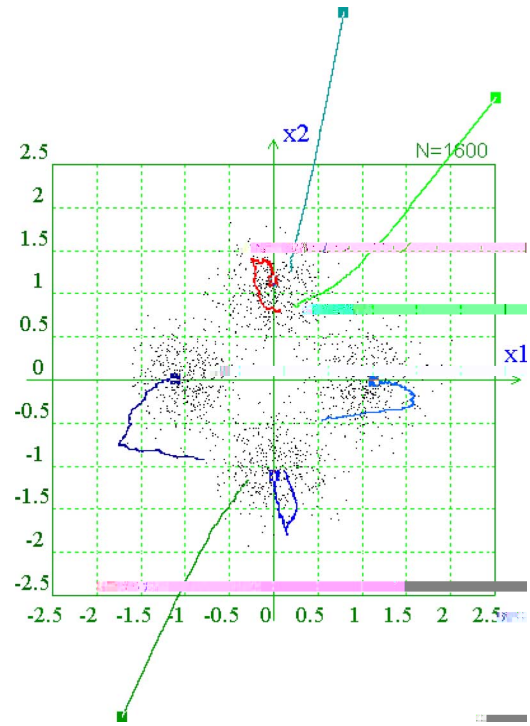
n k 4 η

$\mathcal{S}_1$ $\mathcal{S}_2$

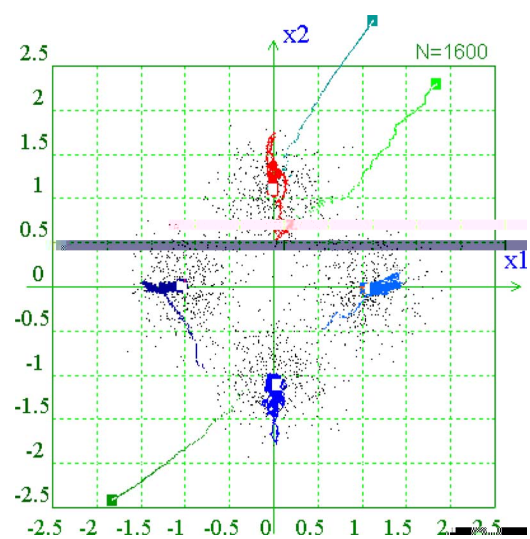
$E(\mathbf{W})$

$\mathcal{S}_2$

$E_2(\mathbf{W})$



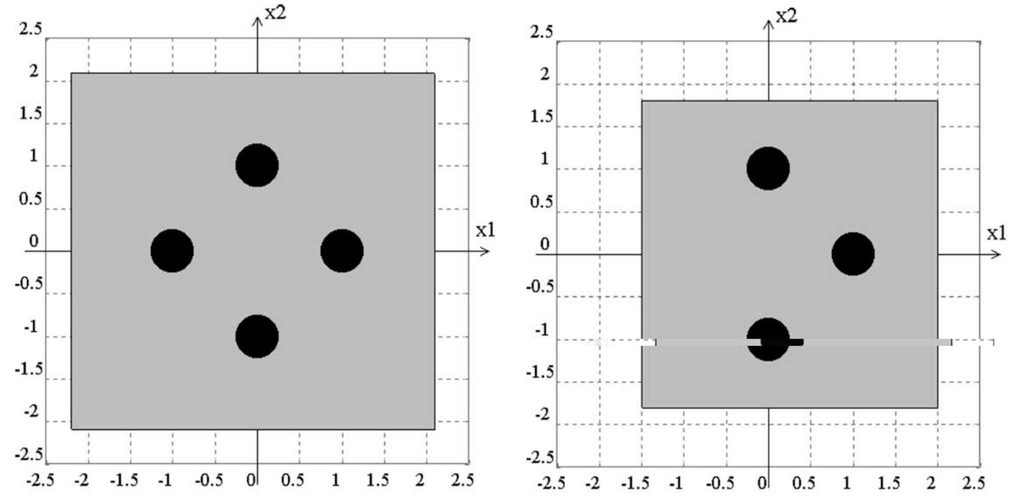
n k 4 η



n k 4 η

$E_2(\mathbf{W})$ $m = [t/5N]$ $[x]$ x t N

$n = 7$ $k = 4$ η/m $m = [t/5]$



(a) (b)

$[-2.1, 2.1] \times [-2.1, 2.1]$

$[-1.5, 2.0] \times [-1.\ , 1.\ ]$

$\Delta E_t(\mathbf{W}) = |E(\mathbf{W}^{(t)}) - E(\mathbf{W}^{(t-1)})| < 10^{-6}$

n

$E(\mathbf{W})$

$E(\mathbf{W})$ n

$n \leq 2k$

3) *A* *C* *(A C) A*:

$\mathcal{S}_3$ $\mathcal{S}_4$ $\mathcal{S}_5$

n k

n

w_{ij} $[a, b]$ $T = 0$

T $\lambda = \exp(-k_1 T - k_0)$ $\eta = \eta_0/(c_1 T + c_0)$ $t = 0$

t

X^μ $\mathcal{S} = \{X^1, X^2, \ldots, X^N\}$

ξ

$\xi > \lambda$

$$\Delta W_i = \begin{cases} \eta(X^\mu - W_i), & i = c(\mu) \\ -\eta\|X^\mu - W_i\|^{-P-2}(X^\mu - W_i), & \end{cases}$$

$\xi \leq \lambda$

$$\Delta W_i = \begin{cases} -\eta(X^\mu - W_i), & i = c(\mu) \\ \eta\|X^\mu - W_i\|^{-P-2}(X^\mu - W_i), & \end{cases}$$

$t < M$ $t = t + 1$

$\lambda < \varepsilon$ $T = T + 1$

$\eta/(T+1)$ $1/T$

M

M n η

ε

10^{-6}

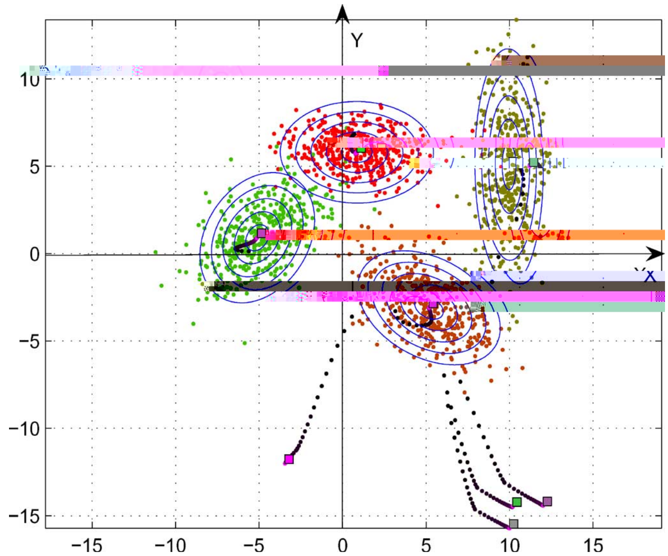
T k_0 k_1 c_0 c_1

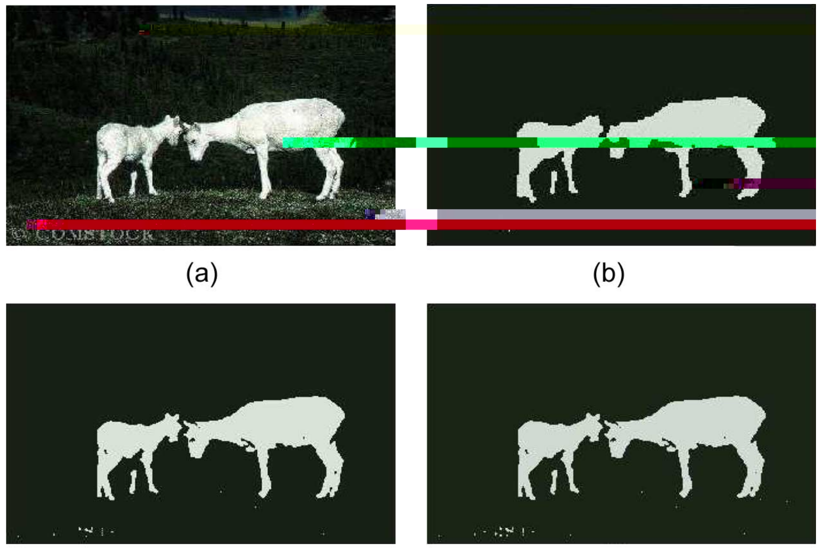
$[a, b]$

$\eta_0 = 0.003$ $M = 100$ $k_1 = 0.005$ $k_0 = 1.200$ $c_1 = 0.015$ $c_0 = 1.000$ $[a, b] = [-1.2, 1.2]$

$T = 10000$

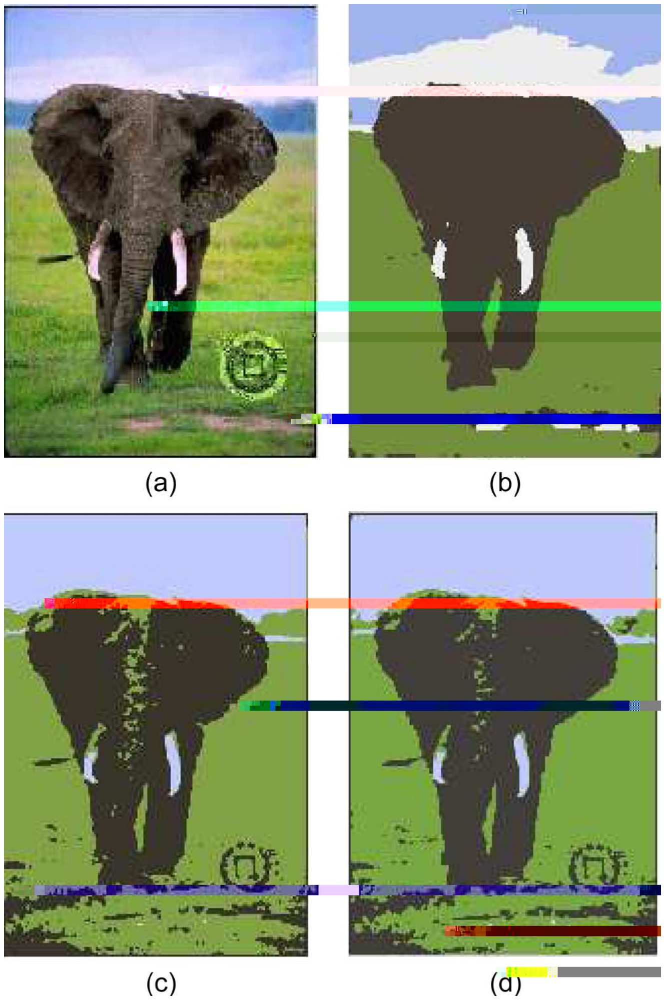
The set of sample data	$\mathcal{S}_1$	$\mathcal{S}_2$	$\mathcal{S}_3$	$\mathcal{S}_4$	$\mathcal{S}_5$
VP = 100%	4-36	4-9	4-8	4-6	3-5
VP $\geq$ 97%	4-62	4-21	4-15	4-9	3-9
VP $\geq$ 95%	4-65	4-38	4-25	4-14	3-10





(a) (b)

(c) (d)



(a) (b) (c) (d)

D. D B D C A

C. C D

$k = 6$

Jinwen Ma

Taijun Wang